THREE ESSAYS ON APPLIED ECONOMETRICS

A Dissertation

by

XIRONG CHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Li Gan
Committee Members,	Guoqiang Tian
	Yuzhe Zhang
	Ren Mu
Head of Department,	Timothy Gronberg

May 2017

Major Subject: Economics

ABSTRACT

In this thesis, I study microeconometrics and its applications in empirical questions. There are three studies, which investigate the housing vacancy rate, sticky demand, and the quantile forecasting methods for time series data.

In the first study, I use a mixture density model to estimate the housing vacancy rate of 15.85% from housing-units level electricity consumption data. This high vacancy rate reflects the high volumes of investment demands in the Chinese housing market.

In the second study, I relax a fundamental assumption in the literature of dynamic pricing, from regular demand to sticky demand and investigate the related economics outcomes.

In the third study, I extend a novel nonparametric quantile forecasting method that is from being only applicable to i.i.d data to being more general applicable that allows for weakly dependent time series data. The simulated and empirical data illustrated the efficiency gain of my method over other existing nonparametric methods.

DEDICATION

This thesis is dedicated to my parents for all of their endless love and continued support. I am extremely grateful and thankful for all they provided me in the past and continue to provide me today. Also, this thesis is dedicated to my advisers, Dr. Li Gan and Dr. Qi Li, who has been great resources of motivation and inspiration.

ACKNOWLEDGMENTS

At first, I would like to thank Dr. Li Gan and Dr. Qi Li for the continuous support of my Ph.D. study and research. Their guidance helped me in all the time of research and writing of this thesis.

Secondly, I would like to thank the rest of my thesis committee: Dr. Guoqiang Tian, Dr. Yuzhe Zhang, and Dr. Ren Mu for their encouragements and comments.

Finally, I thank my fellow Ph.D. classmates in the Department of Economics of Texas A&M University: Darong Dai, Wenzheng Gao, and Zheng Li, for the inspiring discussions and for all the fun we have had in the last five years.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

Econometric methods for empirical questions have been the constant topics in the literature. I develop the novel econometric methods and apply them to answer microeconomics questions. This thesis is organized as following:

In Chapter 1, I propose a novel identification strategy to estimate the housing vacancy rate of 15.85% from housing-units level electricity usage data. Our identification strategy complements the conventional methods when people don't have survey data. This chapter also offers an alternative approach to assess the over-supply risks. Rather than unsold units from newly build housing flows, the households investment housing stocks rise more concerns and have been mostly unexplored in the literature. I offer solid empirical evidence that the strong investment demands existing in the Chinese housing market. These investment properties could hit the market and put more pressure on downward price when housing price decrease since the supply of vacant houses are more elastic than that of occupied ones.

In Chapter 2, The literature on dynamic pricing assumes that consumer demand responds to any changes in price. Recent advances in economics, however, have suggested that consumers may be rationally inattentive and not respond to small price changes, resulting in demand stickiness. We explicitly model the implications of this sticky demand on firm's pricing behaviors. Using a large dataset consisting of eight years of weekly grocery retail data, we estimate the magnitude of demand stickiness and demonstrate how the estimates vary with consumer demographics and product characteristics. Furthermore, we conduct a counterfactual analysis demonstrating the profit improvement a retailer could enjoy by taking into account this demand stickiness when dynamically setting prices to clear inventory.

In Chapter 3, I consider the problem of forecasting a conditional quantile function in a nonparametric framework with time series data. We prove the consistency and asymptotic normality of our nonparametric conditional quantile estimator for absolutely regular processes (β-mixing) data generating processes. We derive the asymptotic distribution of our proposed estimator and conduct Monte Carlo experiments to compare the finite sample performance of our estimator and traditional check function based estimator. Simulation results show that our estimator outperforms the check function based estimator in terms of out-of-sample forecasting. We also apply the estimation method to forecast monthly U.S. housing return based on S&P/Case-Shiller House Price Indices and compare the forecasting results with those obtained by using a commonly used linear conditional quantile model and by using conventional check function based nonparametric conditional quantile estimator. Our estimator forecast well compared to these competitors, especially for data in the tail regions.

2. IDENTIFYING HOUSING VACANCY RATE FROM ELECTRICITY BILLS

The vacancy rate has been the great interest to housing market as being the key factor to assess the supply and demand conditions and related risks on the market. Its importance and related policy implications have been discussed in [1, 2, 3, 4, 5, 6, 7, 8]. Most of the existing works focus on the vacancy rate of rental market in a mature economy, like US and UK. However, not much works can be found that estimate the home owner vacancy rate in the emerging or transitional economies. Among these emerging economies, China has attracted particularly concerns and attentions from scholars and policy makers worldwide. The fast growing housing market of China puts more risks on itself and global economy than that of other mature economies.

Over the last decade, China has experienced an extraordinary successful housing market. Based on National Bureau of Statistics of China (NBSC) data, the real estate industry and the related sectors accounted around 13 percent for national output of China, which has the second largest economy in the world and been the major contributor to the world economic stability since 2008. Also, the housing price maintained real annual growth rate at 17 percent since 2004 [9]. While accompanied with this rapid growth rate of housing market, the average real GDP growth rate is about 10% and average real disposable income growth rate is 9% during that period. As the Chinese economy starts to slow down since 2013, people worry about China would follow the steps of 2008 U.S and 1990s Japan. The crash of the housing booms of U.S and Japan severely damage their domestic economy, which have the contagious effects on global economy. Even more concerns of Chinese housing market are rising after a new wave of price hikes from 2015 to 2016. For example, the housing price of Shenzhen, one of major cities of China, grew 51.9% from Jan, 2015 to Jan, 2016. This persistent faster than economic fundamentals growth also generates a series of social inequality problems. The high income households own multiple houses [10], while, the average families suffer from the huge financial burden from acquiring their homes [11].

The soaring housing price has been attributed to the insufficient supply. The Chinese housing market seems following this logic in the sense that the amount of sold housing areas is increase 10 times from 165.7 millions square meters at 2000 to 1,052 millions square meters at 2014. For easing the fast growing price, the central government planed to build 36 millions welfare housing units nation wide from 2011 to 2015 and actually built around 39 millions housing units. However, even with this kind of effort, the housing price keeps going up. Therefore, the underpinning reasons for the high housing price is beyond the seemingly insufficient supply logic.

It's important and urgent to identify the reasons of this ongoing unhealthy growth and to assess the potential risks of housing market crash in China. The recent news and research papers [12] have showed the growing number of unsold housing properties from newly built housing flows in urban areas. However, the vacant units of housing stocks have been exclusively unexplored in the literature. Due to the nascent nature and data limitation of Chinese housing market, it's challenging to estimate the home owner vacancy rate, which is the leading indicator of the Chinese housing market. The existing works about Chinese housing market vacancy rate draw diverge conclusions on the exact number of vacancy rate. All the existing works are based on the descriptive statistics from survey data. Perhaps, the most widely accepted data set on vacancy rate is the China Household Finance Survey (CHFS) from Southwestern Unviersity of Finance and Economics ([10]). This survey reports the nation wide vacancy rate of 22.4% at 2013 based on the information that the number of housing units each household have from a sample of 28,000 households under the presumption that each household can not occupy more than one housing unit simultaneously. However, [13] reports a much lower vacancy rate of 5.3% at 2009 based on the housing units owning and occupying information from Urban Household Survey data from NBSC.

All the existing works estimate the vacancy rate from survey data. However, as commonly acknowledge, the survey data demands more resources to collect than the transaction data and being rare for developing countries. In this paper, we propose an alternative identification strategy, which complements the conventional vacancy rate estimation method and offers an option when people don't have the survey data. The data set consists of 1.2 million housing units electricity usage information from one of the most populated and economically important cities of China. Under the assumption that vacant housing units should have zero or low electricity consumption, we propose a mixture density model to estimate the housing vacancy rate. First, we focus on the non-zero electricity consumption households. We find out that there are two peaks in the non-zero consumption households PDF graph, one peak at very low electricity consumption, 73 kilowatts-hour/year, and another one at household normal electricity consumption, 1558 kilowatts-hour/year. Among the non-zero electricity consumption housing units, We assume that one type housing units with low electricity consumption are occasionally visited by owner and be vacant for most of the time, and another type housing units with normal electricity consumption are dwelling units that are occupied by households. Thus, each of the housing units types have their own underlying electricity usage distributions and a mixture density model could be applied to identify each of them. Second, with 8.27% zero-electricity consumption of total sample observations, we find the percentage of housing units with low electricity consumption is 7.15%. After re-scale, the vacancy rate is 15.85%, which is lower than the CHFS survey reported nation wide 22.4% at 2013, but much higher than the [13] reported 5.3% at 2009. Since our results are based on one of the most populated cities, which have the highest housing price and are widely assumed to have the lowest vacancy rate in the nation, a higher vacancy rate is expected on national level. We should also note that the home owner vacancy rate was 2.5% at United States even during the peak of 2007 U.S housing boom based on American Housing Survey.

This vacancy rate indicates a massive part of Chinese households purchasing housing units as investment vehicles and leave them vacant, which reflect the high investment demands [14] on the Chinese housing market. The fast growing housing price generates tremendous amounts of housing market capital gains, which stimulated the investment motives on the market and being maintained by the household's high expectations of future income and economy growth. The same phenomenon, homeowners momentum, also happened during the 2002-2006 U.S housing boom. [15] empirically find out that more people would like to own a house as housing price keeps going up. Our estimated vacancy rate offers important economic insights for identifying the risks of Chinese housing market crash. Under the assumption that the vacant housing units are more elastic than occupied ones, the households are very likely to sell their vacant houses on secondary market once housing price decrease, which generate a serious risks of housing market crash. Given the fact that the Chinese central government planed massive welfare housing construction project since 2011, this risk could be even amplified as the economy growth slows down.

The policy implication of this paper helps reconciling the debate on whether people should levy the property tax in China to ease the fast growing housing price. Even though there has been constant debate over the degree of property taxes capitalization, the public finance literature widely accepted that the property taxes can be partially capitalized into home values [16, 17, 18, 19]. In more recent studies, [20] confirms that tax on vacant units reduces the vacancy in French rental housing market. Intuitively, the housing values, like any other asset values, should equal to the present value of housing flows from owning it. The property tax will increase the housing holding cost, which lead to a lower home values. As many of the existing studies showed, the property tax can effectively reduce the housing capital gains. For easing the identified investment demands and soaring housing price, we suggest that China should levy the property tax.

2.1 Institutional Background

The Chinese housing market is still nascent as the houses were not commercialized until 1998. Before then, the residential properties were allocated by the employment relationship between state-owned working units and employees. The working units assign various sizes and locations of housing units to their employees depending on the length of employment and size of households, among other factors, as parts of employment benefit package. In 1998, with the progress of economic reform and privatization since 1978, the traditional housing allocation mechanism was terminated and the "commodity houses" and "welfare houses" were introduced. Similar to western countries housing systems, the price of "commodity houses" is determined by the market and the "welfare houses" are supposed to be assigned to low or medium income households. Since then, the housing price start to soar with both the strong demand and strong supply of residential housing.

In order to understand the supply and demand metrics better, we firstly look at the economic fundamentals in China. As well known, China has experienced an spectacular economic growth in recent decades with an average annual real GDP growth rate about 10 percent. However, we should note that there are two drops in GDP growth as shown in figure 2.1. The first one is a sharp drop happened between 2008 and 2009 mainly because of the 2008 financial recession that originated from U.S and spread world wide rapidly. For preventing the economy sharp drop from this recession, the Chinese central government initiated a timely comprehensive 580 billions U.S dollars stimulus package, which invested a substantial proportion to the real estate sector. The more concerning drop happened at 2013 as the slowest economic growth rate of 7.1 in recent 20 years, which makes the public concern the crash of Chinese housing market boom.

Accompanied with economic growth, the urbanization process in China experienced a steady growth. The urban population grows from 460 millions at 2000 to 750 millions at 2014 with average growth rate of 3.56.

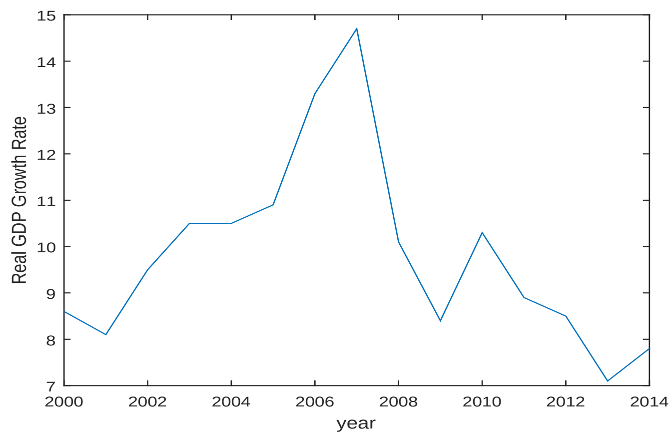

Figure 2.1: GDP Growth

The newly built housing areas grows from 244 millions square meters at 2000 ti 1248 millions square meters at 2014 with average annual growth rate of 13.2%. As shown on the figure 2.2, the housing supply keeps increasing even during the 2008 recession time. We think that is mainly because of the comprehensive stimulus package during that time. While, the supply start to fluctuate, but stays high, since the Chinese economy slowing down of 2012. In addition, targeting on continuously rising price and inequality problems, the central government launched the welfare housing program to increases the housing supply. The most remarkable welfare housing project was initiated by the he Ministry of Housing and Urban-Rural Development at 2011 to offer 36 million welfare housing units nation wide within 5 years. According to the recent government reports, the actual number of completed welfare housing units is 39.7 million by the end of 2015.

The sold housing areas grow from 165 millions square meters at 2000 ti 1052 millions

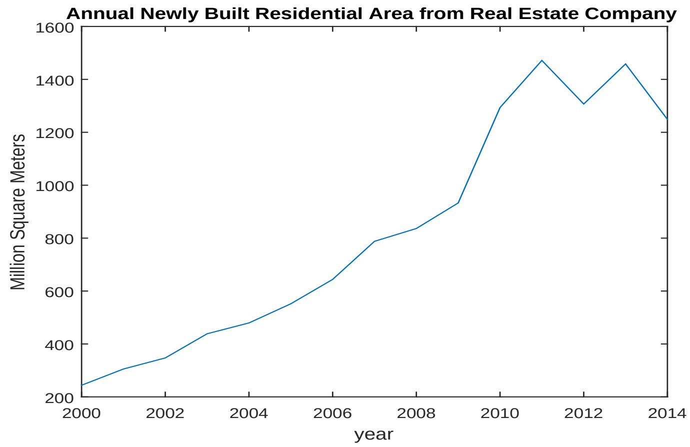

Figure 2.2: New Build

square meters at 2014 with average annual growth rate of 15.4%. As shown in the figure 2.3, there are only two decreasing episodes during the period of study, which coincide with two drops of GDP growth. First demand decreasing happened during 2008 financial recession time, but recovered soon and increased faster than before as the central government stimulus package took effect. Another decreasing happened at 2014 as two years after the Chinese economy starts to slow down.

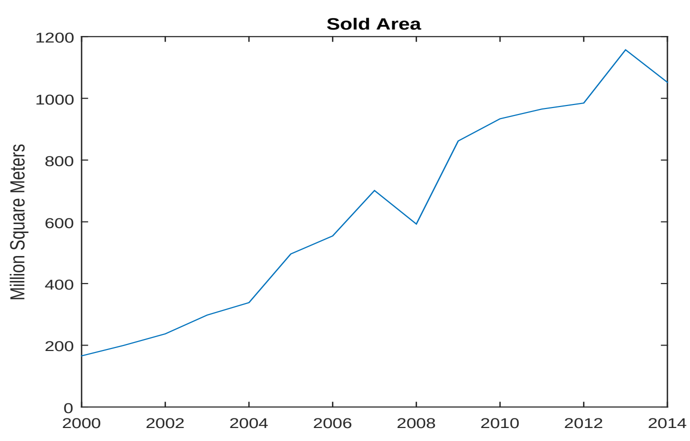

Figure 2.3: Annual Demand

The property tax, as one of the primary tax sources for local governments in many developed countries, has not been levied in China yet. As an alternative tax source, the land sale revenue largely support the Chinese local governments for public goods provisions given all the land is owned by government in China. Therefore, the local governments have incentives to sell the land for a higher price, and these incentives transfer to housing prices as well. Given the decade long Chinese housing market booming, there is a heated nationwide debate that whether the central government should levy the property tax for easing the soaring housing price . The pro side argues that the property tax can relive the local government from land sales and con side argues that the property tax could crowd out the household consumption [21]. In order to preparing for the potential tax reform, the government initiated the pilot property tax trials in January 2011. These trails were implemented in Shanghai and Chongqing, where are two important and representative cities of east side and west side of China. They provide good policy experimental results for analyzing the impact of property tax on housing prices. [22] and [23] find out the negative relationship between the incidence of property tax trials and the housing price.

In Shanghai, the property tax was targeting at multiple housing units owners. For households with Shanghai municipal residential registration (hukou), they pay 0.6% tax rate for housing properties other than their primary dwelling housing unit in general. The tax rate can reduce to 0.4% if their housing properties are cheaper than twice the average housing price of residential homes at city level. The tax exemption is 60 square meters per household member. In Chongqing, the property tax was targeting at luxury housing properties. The tax rate is 0.5% for housing properties with price lower than around 2000 U.S dollars per square meter and increase to anywhere between 1% and 1.2% for housing properties with price higher than that. The tax exemption is 180 square meters for independent house and 100 square meters for apartments.

2.2 Model and Estimation

We build a mixture density model to identify the different types of housing usages from a set of electricity bill data such that the housing vacancy rate could be recovered.

Let x denote the 2014 electricity bill of an apartment from this major city in China. There are three different cases to generalize the housing situations:

(i) the apartment is fully occupied so that x is a draw from an unknown distribution with a positive housing electricity consumption.

(ii) x may be a very small number suggesting that this apartment is basically vacant in 2014, the owner may occasionally visit the apartment, say, for a regular weekly or monthly check.

(iii) x may be zero meaning that this apartment was completely vacant during 2014 (sold un-decorated apartments exist in the Chinese housing market so that no one can possibly live there).

We use H_1, H_2 and H_0 to denote the numbers of housing units for cases (i), (ii) and (iii), respectively, then the housing vacancy rate for the city is defined as

$$VR = \frac{H_2 + H_0}{H_1 + H_2 + H_0}$$

The information on H_0 is available from the data. The difficulty is to identify H_2 from $H_1 + H_2$. Let x_1 and x_2 be random draws of electricity consumption from type (i) and type (ii) apartments, and $F_1(\cdot)$ and $F_2(\cdot)$ be their cumulative distribution functions with probability density functions $f_1(\cdot)$ and $f_2(\cdot)$, respectively. Conditional on $x > 0$, we use p to denote the percentage of apartments that belong to type (i), and $1 - p$ denote the percentage of apartments that belong to type (ii). Also, let p_0 denote the percentage of apartments from type (iii).

To identify/estimate p, we propose a mixture density for positive electricity consumption ($x > 0$).

$$f(x|\theta) = p\ f_1(x|\theta) + (1-p)\ f_2(x|\theta) \tag{2.1}$$

The parameter of interest p and other relevant parameters θ will be estimated by maximizing the following summation of log likelihood function:

$$L = \sum_{i=1}^{N} \ell_i(\hat{\theta}) \tag{2.2}$$

where $\ell_i(\hat{\theta}) = log f(x_i|\hat{\theta})$

Denote total number of housing units $H = H_0 + H_1 + H_2$, We have $p_0 = \frac{H_0}{H}$, $p = \frac{H_1}{H_1+H_2}$, $1-p = \frac{H_2}{H_1+H_2}$. Since both p and $1-p$ are conditional on $X > 0$, we need to re-scale them with factor $\frac{H_1+H_2}{H}$ to compute the vacancy rate VR. We denote probability of being type (ii) households as $p_2 = (1-p)\frac{H_1+H_2}{H} = \frac{H_2}{H_1+H_2}\frac{H_1+H_2}{H} = \frac{H_2}{H}$. Then, we have

$$VR = p_0 + p_2 \tag{2.3}$$

2.3 Data and Empirical Results

We have data of electricity bills (in 2014) about 1.2 million units of apartments, which are already sold to households and registered in electricity companies. We trim out data with $X > 10,000$ (kilowatts-hour). There are several reasons that one may want to trim out extreme tail observations. First, very large electricity consumption units could be super luxury residential units, which are not the interest of this paper. Second, extreme tail observation can have influential impact on estimation results. Removing data with $x > 10,000$ drops 4726 observations, which are less than 0.5% of the total observations. In addition, there are $103,546$ observations with $x = 0$, which is about 8.7% of the sample. After remove data for both $x = 0$ and for $x > 10,000$, we obtain a sample $n = 1,090,898$.

The raw data histogram is plotted in figure 2.4.

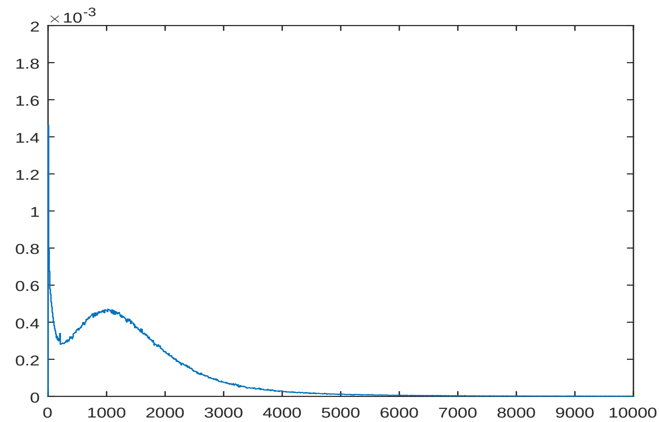

Figure 2.4: Raw Data Histogram

We start with a mixture of normal distribution model for benchmark estimation purpose. The normal distribution is the most popular distribution assumption in the literature and the two peaks in figure 2.4 seem to could be captured by a mixture of two normal distribution.

A Normal density function has the following form.

$$n(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad 0 < x < \infty. \tag{2.4}$$

We assume that a random draw from positive x, it is a draw from $n(x|\mu_1, \sigma_1^2)$ with probability p, and from $n(x|\mu_2, \sigma_2^2)$ with probability $1 - p$, Therefore, the mixture normal density function for x is given by

$$f_n(x) = p\ n(x|\mu_1, \sigma_1^2) + (1 - p)\, n(x|\mu_2, \sigma_2^2), \tag{2.5}$$

where $n(x|\mu_1, \sigma_1^2)$ is defined in (2.4). We estimate the parameters p, μ_j, σ_j for $j = 1, 2$ by the maximum likelihood method and the results is presented in Table 2.1.

Table 2.1: Two Normal Mixed Estimation Results

Parameter	p	$1-p$	μ_1	σ_1^2	μ_2	σ_2^2
Values	0.15	0.85	$3.16{*}10^3$	$1.7{*}10^3$	$1.13{*}10^3$	715

We plot the mixture of normal density with larger mean in figure 2.5, with smaller mean in figure 2.6, and the entire mixture of two normal densities in 2.7. We compare the mixture of two normal densities with raw data histogram in figure 2.8. We found out that the two normal density mixture can only capture one of the peaks, which is the one with larger mean value and larger variance picking up the fat tail of x.

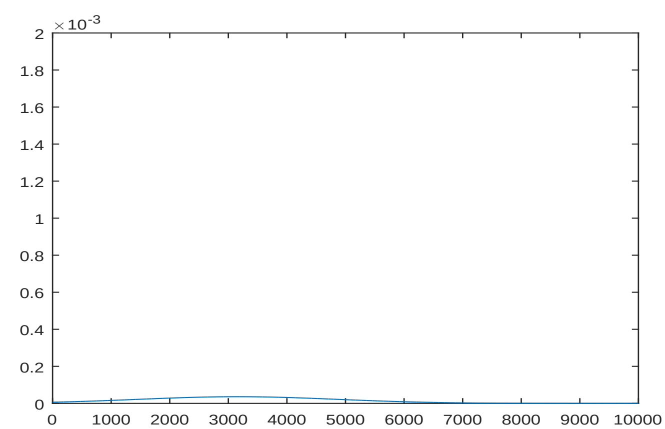

Figure 2.5: The Normal Density Mixture P.D.F with Larger Mean

Since the Normal density mixtures can not capture the observations, we propose a mixture of Gamma density to capture the shape of raw data. A Gamma density function

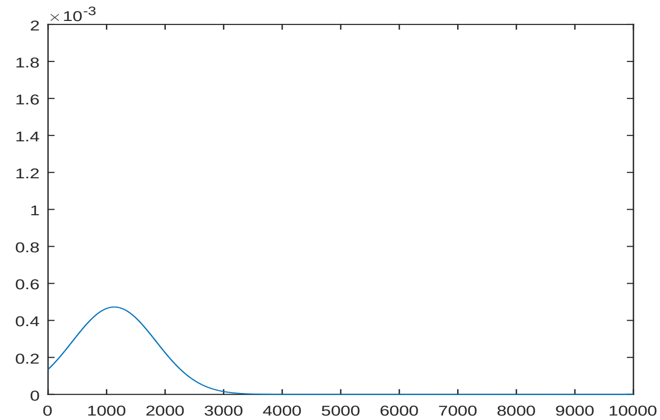

Figure 2.6: The Normal Density Mixture P.D.F with Smaller Mean

has the following form:

$$g(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \qquad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0. \tag{2.6}$$

There are several reasons for using Gamma densities to describe electricity consumption. First, $x > 0$ is satisfied as there cannot be negative electricity consumption. Second, when $0 < \alpha < 1$, the gamma density function has a large peak at $x = 0$, which can be used to describe electricity consumption for type (ii) houses; when $\alpha > 1$, the Gamma density function takes value 0 at $x = 0$ and it is peaked at $x = (\alpha - 1)/\beta > 0$, which can be used to model electricity consumption for "non-vacant or occupied houses". Therefore, Gamma distribution has a much longer fat tail than normal distribution does.

We assume that a random draw from positive x, it is a draw from $g(x, \alpha_1, \beta_1)$ with probability p, and from $g(x, \alpha_2, \beta_2)$ with probability $1 - p$. Thus, the mixture of gamma density function for x is given by

$$f_g(x) = p\; g(x, \alpha_1, \beta_1) + (1 - p)\, g(x, \alpha_2, \beta_2), \tag{2.7}$$

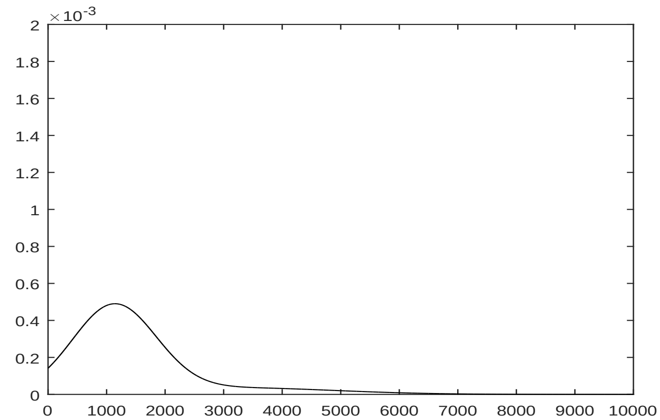

Figure 2.7: Two Normal Densities Mixture

where the Gamma density function $g(x, \alpha, \beta)$ is defined in (2.6). We estimate the parameters p, α_j, β_j for $j = 1, 2$ by the maximum likelihood method.

We present the estimation results in table 2.3, plot the estimated p.d.f and compare it with raw data histogram in figure 2.9. We observe that the estimated Gamma density traces the histogram's shape quite well.

We know that if $x \sim$ Gamma-distribution with pdf $f_g(x, \alpha, \beta)$, then $E(x) = \alpha\beta$ and $Std(\beta) = \sqrt{\alpha\beta^2}$. Using the estimated parameters, we see that $E(x_1) = (2.05)(760) = 1558$ and $Std(x_1) = 1088$; while $E(x_2) = 73.0$ and $Std(x_2) = 82.4$. The average electricity consumption for type (ii) houses is less than 5% of that of type (i) houses. Indicating that these houses were only occasionally visited by their owners and therefore are very likely to be 'vacant houses'.

Table 2.2: Two Gamma Mixed Estimation Results

Parameter	p	$1-p$	α_1	β_1	α_2	β_2	μ_1	σ_1^2	μ_2	σ_2^2
Values	0.922	0.078	2.05	760.0	0.785	93.0	1558	1088	73	82.4

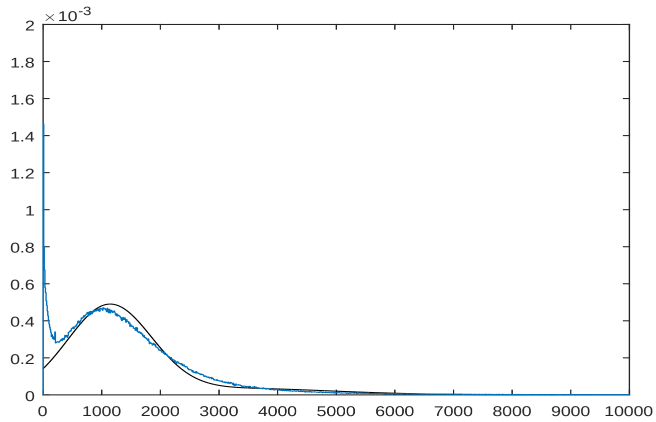

Figure 2.8: Two Normal Densities Mixture and Raw Data Comparison

For model selection, we use the [24] method. We conduct the non-nested likelihood ratio test on the mixture of normal model and the mixture of Gamma model. The null hypothesis is that the likelihood of the mixture of Gamma distributions equals to that of the mixture of normal distribution.

Let $f_g(x|\theta_1)$ and $f_n(x|\theta_2)$ be competing models for the density of $D(x_i)$, where both may be misspecified. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the QMLEs converging to θ_1^* and θ_2^*. Let $L_m = \sum_{i=1}^{N} \ell_{im}(\hat{\theta}_m)$ be the quasi-log likelihood evaluated at the relevant estimate for $m = n, g$. Then

$$\begin{aligned} N^{-1/2}(L_1 - L_2) &= N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\hat{\theta}_1) - \ell_{i2}(\hat{\theta}_2)] \\ &= N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\theta_1^*) - \ell_{i2}(\theta_2^*)] + o_p(1) \end{aligned} \tag{2.8}$$

Under the null hypothesis $H_0 : E[\ell_{i1}(\theta_1^*)] = E[\ell_{i2}^*(\theta_2^*)]$,

$$N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\hat{\theta}_1) - \ell_{i2}(\hat{\theta}_2)] \to N(0, \eta^2) \tag{2.9}$$

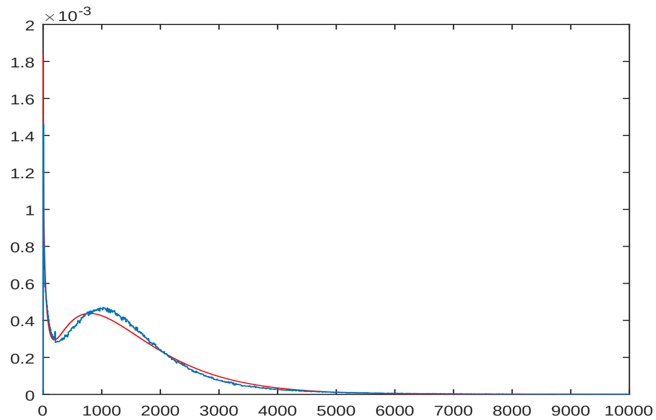

Figure 2.9: Two Gamma Densities Mixture and Raw Data Comparison

where $\eta^2 = var(\ell_{i1}(\theta_1^*) - \ell_{i2}(\theta_2^*))$. A consistent estimator of η^2 is

$$\hat{\eta}^2 \equiv N^{-1} \sum_{i=1}^{N} [\ell_{i1}(\hat{\theta}_1) - \ell_{i2}(\hat{\theta}_2)]^2 \tag{2.10}$$

Vuong's model non-nested test statistic is

$$\begin{aligned} VMS &= N^{-1/2}(L_1 - L_2)/\hat{\eta} \\ &= \frac{N^{-\frac{1}{2}} \sum_{i=1}^{N} [\ell_{i1}(\hat{\theta}_1) - \ell_{i2}(\hat{\theta}_2)]}{\{N^{-1} \sum_{i=1}^{N} [\ell_{i1}(\hat{\theta}_1) - \ell_{i2}(\hat{\theta}_2)]^2\}^{1/2}} \to N(0,1) \end{aligned} \tag{2.11}$$

under H_0.

A simple way to obtain a valid test is to define $\hat{d}_i = \ell_{i1}(\hat{\theta}_1) - \ell_{i2}(\hat{\theta}_2)$ for each i and then simply regress $\hat{d}_i$ on unit constant to test if the estimated coefficient is significantly different from zero.

We present the model selection testing results in table 2.3 and find out that the coefficient is positive and significant. Thus, the H_0 is rejected and we conclude that a mixture Gamma density fits the data significantly better than a mixture normal density.

Table 2.3: Model Selection Test Statistics

Variable	Coefficient	T-Statistics	P-Value
Constant	0.11	241.96	0

The gamma mixture model estimation result suggests $1-p = 0.0778$ and $p_2 = 0.078 * \frac{1{,}090{,}898}{1189718} = 0.0715$. Thus, about 7.15% of housing units belong to type (ii) whose electricity consumption is extremely small compared with that of "non-vacant houses". This, plus ($p_0 = 8.27\%$) gives us an estimated vacancy rate $p_0 + p_2 = 0.0827 + 0.0715 = 15.35\%$ for the city, which is close to the 2013 SWUFE survey of 19% vacancy rate for this city.

Next, we do the same analysis as above but considering the location of apartments. We consider three cases: (a): inner circle houses: an apartment is close to the center of the city (b) middle circle houses (c) out circle houses.

The location (distance to city center) variable is missing for 451,946 data points (close to 40% sample with $x > 0$). For the remaining data with location variable, we split the sample into the above three cases according to their locations. The estimation results are given in Table 2.3. From Table 2.3 we observe that the vacancy rates (conditional on $x > 0$) varies with location. For apartments located at the inner circle, it has the lowest vacancy rate of 9.36%, followed by 15% for apartments located at the out circle. For apartments in the middle circle, it has a relatively large vacancy rate of 16.8%. This result is reasonable and can be explained as follows. First, for apartments close to the city center, their values as well as rents are much higher than apartments outside this inner circle. It is too costly to let these high value, high demand apartments to be vacant. For multiple apartments families, if they do not live at an apartment located at the inner circle, it is more likely for them to rent out the apartment, resulting in a low vacancy rate. For the middle circle apartments, they have high investment values and at the same time, they are not as expensive as apartments located at the inner circle. Therefore, many people would like to own apartments located in the middle circle as investments. Also, the rent as well as rental demand in this circle is not as high as those in the inner circle, making it less attractive for owners to rent them. This results in a high vacancy rate for apartments located at the middle circle. Finally, for apartments located in the out circle, given their long distance to city center and that many accompany facilities (such as shopping centers, schools, etc.) are not well established yet, their investment values is much lower than apartments in the middle circle. This makes people less likely to purchase multiple apartments in the out circle as investments, leading to its lower vacancy rate than that for the middle circle.

Table 2.4: Parameter Estimation V.S Location

location	sample size	p	$1-p$	α_1	β_1	α_2	β_2
inner circle	174,030	0.944	0.056	2.309	673.6	0.858	111.7
middle circle	156,672	0.9313	0.0687	1.8731	843.8	0.9207	97.67
out circle	411,796	0.9235	0.0765	2.5628	579	0.7	122

Table 2.5: Occupancy/Vacancy Rate V.S location

location	Size of Zeros	p_0	p_2	VR
close distance	6,932	0.0398	0.0538	0.0936
middle distance	16,683	0.1065	0.0614	0.1679
far distance	32,176	0.0794	0.07	0.1494

The histograms and estimated mixture Gamma densities for inner circle, middle circle and out circle are given in figures 2.10 to 2.12.

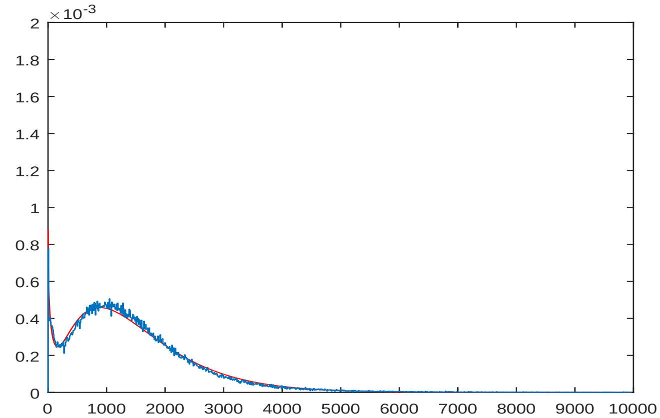

Figure 2.10: Histogram for Inner Circle

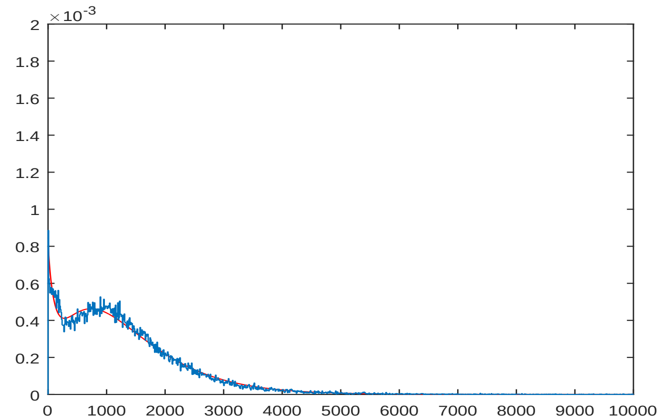

Figure 2.11: Gamma Mixture Density for Inner Circle

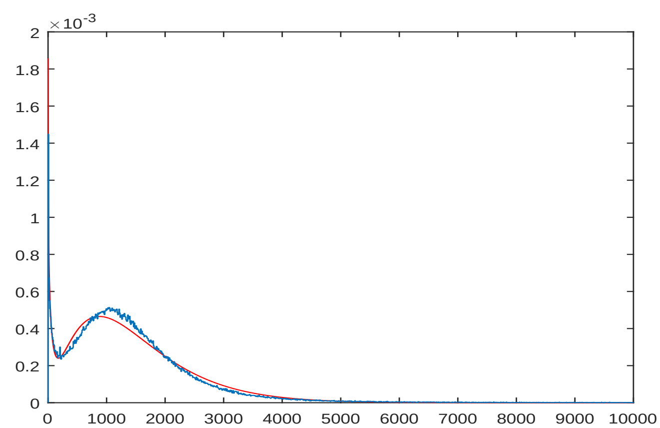

Figure 2.12: Histogram for Middle Circle

3. RETAILER STRATEGIC PRICING UNDER STICKY DEMAND

With recent advancement in the availability of transactional data and the computing power to process these data, retailers become increasingly sophisticated in setting prices to match the supply and demand. Hence, dynamic pricing, whereby retailers adjust prices according to demand and supply information to boost revenue and profit, becomes an active research area in recent years [25, 26, 27, 28]. One of the foundational assumptions in dynamic pricing is that consumers respond to price changes instantaneously. However, as consumers we may not always respond to small price changes. For example, we may not buy more potato chips simply because its price is reduced by a few pennies. This, at the aggregated level, may result in demand stickiness; that is, demand may not shift, as researchers of dynamic pricing would have assumed, to small price changes. This is consistent with the notion of rational inattention [29, 30]. Inspired by prior theoretical work, we aim to empirically identify and estimate the magnitude of demand stickiness using longitudinal data from a grocery chain, with implications for dynamic pricing.

While macroeconomists have found that the market does not immediately respond to market shocks such as new monetary policies, psychologists and marketing researchers have similarly found that consumers do not change their consumption or saving plans after experiencing small income shocks. In a retail environment the existence of these rationally inattentive consumers implies delays or lack of actions to small price changes by these consumers. That is, there exists demand stickiness due to consumer's rational inattention. At the same time, the notion of rational inattention also implies that consumers will re-optimize when facing big price changes.

Clearly, these provide opportunities for retailers to exploit inattentive consumers. To be specific, it is beneficial for a retailer to raise prices in small increments so consumers will not reduce their purchasing quantities. Conversely, a retailer is more likely to decrease prices in larger increments so that consumers do increase their purchasing quantities. In other words, if we observe the pricing behaviors of a retailer that exploits inattentive consumers, we should see many more small price increases than decreases. Indeed, [31] documented such an asymmetric price changing pattern using eight years' of weekly Dominick's data [1] by comparing the number of small price increases and decreases. However, to make [31]'s results meaningful and useful to operations management researchers and practitioners, one needs to estimate the magnitude of the consumer inattentive range, i.e., the price range within which consumers do not re-optimize their purchase quantities. This is the focus of the current paper.

In this paper, using the same data as [31], we estimate the magnitude of this consumer inattentive range econometrically using a structural model. We assume that the retailer's behaviors are optimal given consumer inattention, and we estimate the consumer inattentive range from the retailer's best response function. To check the robustness of our results, we also directly estimate the consumer inattentive range based on the observed sales quantities when prices change. Both models generate comparable results.[2]

In addition, to explore the underlying determinants of sticky demand, we investigate the relationship of the magnitude of demand stickiness with product and consumer characteristics. We find that the consumer inattentive range is positively correlated with the proportion of hurried shoppers, and negatively correlated with the average household size and consumer education level. Besides, consumer inattentive range is similar between private label and national brand products and is independent of the intensity of competition. Finally, we conduct a counterfactual analysis to show that a retailer can increase its revenue by about 1% if it considers demand stickiness when setting its prices dynamically to

[1] Dominick's was a grocery chain in the greater Chicago area. The data span from 1989 to 1997.

[2] Throughout the paper, we use "consumer inattentive range" and "magnitude of demand stickiness" interchangeably.

clear inventory. These results help illustrate the managerial relevance of studying demand stickiness.

To the best of our knowledge, this paper is among the first to empirically estimate the magnitude of demand stickiness in a grocery retail environment. Our findings have important theoretical and practical implications regarding how to better leverage dynamic pricing for higher profits. This paper also contributes to the growing literature concerning the effect of rational inattention on consumption behaviors.

The rest of the paper is organized as follows. In section 3.1, we review pertinent literature. In section 3.2, we describe the theoretical model that derives the retailer's optimal price with inattentive consumers. In section 3.3, we propose econometric models and estimate the rational inattentive range as well as the relationships of this range with product characteristics and consumer demographics. We report the simulation results of the counterfactual analysis in section 3.4.

3.1 Literature Review

Retailer's strategic pricing behavior under sticky demand is related to dynamic pricing, retail pricing and psychological/economic theories for rational inattention. We elaborate on the most relevant research in each area below.

The literature on dynamic pricing has evolved from a focus on single product, monopolist, and perfect information settings to those that include multi-products, competition, and limited demand information [32, 28]. Despite the progress made in this literature with less restrictive assumptions and more realism, very few papers have studied behavioral effects on dynamic pricing [33]. [34] study the effect of reference price on purchase decisions, and develop a model based on the notion that consumers have memory and their decisions could be affected by cognitive limitations. [35] embeds consumer inertia into a decision model to reflect the phenomenon that consumers may postpone their purchase decisions even if it is the optimal timing to purchase. He finds the dual effects of inertia on firm profit such that the inertia can decrease demand at beginning and increase desires to purchase later. [36] consider the effects of consumers regret when they make purchase decisions. Regret can induce consumers to buy either later (yielding inertia) or earlier than the optimal purchase timing. [37] study how regret and availability misperceptions motivate consumer purchase decisions, and argue that ignoring these effects could significantly hurt firm operations and profit. Dynamic pricing has been applied to many other settings, such as inventory problems [25], strategic consumer problems [38, 39, 26, 27], and the retail industry [40], among others.

In the area of retail pricing, [41] builds a structural model to investigate firm's cost of price adjustment mechanism and to estimate the magnitude of adjustment cost using weekly retail pricing data from a grocery chain. Another stream of works has focused on the effect of timing, such as holidays and weekends, on retail pricing patterns [42, 43, 44]. The determinants of retail pricing decisions has been studied as well. For example, [45] use a simultaneous equation model to study the factors that affect retailer pricing decisions. In similar spirits, [46] uses time-series data to study the effects of price history, wholesale prices and demand on retail prices.

While research on dynamic pricing offers the insight that consumers face cognitive limitations and their limited cognition could affect retail pricing, there is little empirical work in this area. Research on retail pricing, though largely empirical, has not studied the behavioral effects of consumer's cognitive limitations on pricing [47]. Our paper fills this gap in the literature by assuming that consumers are rationally inattentive and cannot fully respond to a small price change, and we estimate the ensuing demand stickiness due to consumer's rational inattention. This assumption is supported by research in economics and marketing on rational inattention [48, 29, 30].

Most relevant to our current work, two models of rational inattention have been proposed. In one model, [30] borrows Shannon's information theory and models a utility-maximizing agent as an information processing channel. Due to channel capacity constraints, the agent chooses not to pay attention to all of the information. As a result, as Sims observes, "[a]gents with rational inattention ... will respond with more delay and information-processing error – or may not respond at all – to fluctuations that are small and therefore relatively unimportant to them." [29] proposes an alternative model in which the costs of monitoring, absorbing and processing information force utility-maximizing agents to update their choice of actions only sporadically. Therefore, "[t]hese consumers rationally choose to only sporadically update their information and re-compute their optimal consumption plans. In between updating dates, they remain inattentive." Regardless of which model one adopts, in a retail environment the existence of inattentive consumers implies delays or lack of actions to a small price changes on the consumers' side. That is, there exists demand stickiness among inattentive consumers. On the other hand, both models predict that consumers will be attentive to big price changes.

3.2 The Model

In this section, we model the retailer's profit maximization process. Assume there exists menu cost and it is a fixed cost [49, 50]. The menu cost is incurred when the retailer decides to change the price; that is, when the current period's price is different from last period's. The retailer maximizes its profit by choosing a retail price in each time period. Assume that the demand function and the cost of the products are exogenous. The decisions for the retailer to make are: 1. Whether or not to pay the menu cost to change the price; and 2. if the menu cost is paid, what is the new optimal price to set? The first question is answered through backward induction. If the new optimal price in the second stage yields an excess profit (i.e., relative to the profit generated by the old price) that is greater than the menu cost, the retailer chooses to pay the menu cost. Otherwise, the retailer chooses not to change the price. Note that the menu cost becomes sunk in the second stage. In the following, we focus on the second stage, which involves the modeling of the demand function structure.

3.2.1 The General Framework

For analytical trackability, we assume that for a given product, its demand function is determined only by its own historical and present prices[3] but not by other products' prices or its own future prices. Thus, we write $D(p_t)$ as the demand function of a given product, where p_t is the price of this product at time t. Note that this demand function takes only current period's price as the argument. It does not account for any reference effect. We call this the simple demand function, whereby every consumer re-optimizes her purchasing behavior according to the latest price. Following [29] and [30], we incorporate consumer's rational inattention by assuming that only some of the consumers will re-optimize in response to the latest price. Denote the fraction of consumers whose last re-optimization happens at time t as w_t. Denote $\tilde{D}(p_t, p_{t-1}, \cdots, p_1)$ as the effective demand given by

$$\tilde{D}(p_t, p_{t-1}, \cdots, p_1) = w_t D(p_t) + w_{t-1} D(p_{t-1}) + \cdots + w_1 D(p_1) \tag{3.1}$$

The effective demand function is a weighted sum of all simple demand functions from time 1 to time t. For analytical trackability, we simplify the effective demand function by doing the following approximation. Since each consumer re-optimizes sporadically, for those time periods far away from time t, the fraction w_t should be close to zero. This enables us to ignore those time periods before $t-1$ without causing significant imprecision. The

[3] In other words, we assume consumers are myopic. Incorporating forward-looking into our model should not change our results qualitatively.

simplified version of the effective demand function is given by

$$\tilde{D}(p_t, p_{t-1}) = w_t D(p_t) + w_{t-1} D(p_{t-1}) \tag{3.2}$$

Since w_t and w_{t-1} should cover all of the consumers, we have $w_{t-1} = 1 - w_t$. The above equation can be further rewritten as

$$\tilde{D}(p_t, p_{t-1}) = w_t D(p_t) + (1 - w_t) D(p_{t-1}) \tag{3.3}$$

After specifying the effective demand function, we bring in cost to construct the profit. The cost is assumed to be exogenous and denoted as c_t. Denote $\pi(p_t|p_{t-1})$ as the profit a retailer can make at time t, which is given by

$$\pi(p_t|p_{t-1}) = \tilde{D}(p_t, p_{t-1}) \times (p_t - c_t) \tag{3.4}$$

The profit is conditional on the last period's price because the effective demand function takes last period's price as an argument. The retailer chooses the retail price p_t to maximize its profit $\pi(p_t|p_{t-1})$. Now we have a general idea of the profit maximization process. In the next section, we will discuss the detail of the simple demand function and the fraction of consumers who re-optimizes at time t, to derive an analytical solution for the retailer's profit maximization problem.

3.2.2 The Analytical Solution

To obtain the analytical solution to the retailer's profit maximization problem, we first make more assumptions about the simple demand function. Assume that the simple demand function is linear in retail price, i.e., $D(p_t) = \beta_0 + \beta_1 p_t$, where $\beta_0 > 0$, $\beta_1 < 0$ and $p_t \leq -\beta_0/\beta_1$. To avoid unnecessary complexity, we assume that the cost c_t is low enough for the retailer to a make positive profit, i.e., $D(c_t) > 0$ for all t. The next question is how to determine the fraction of consumers who re-optimizes at time t. We model the re-optimizing process by introducing a re-optimizing threshold, denoted as L. The threshold is comparable to the price change. A consumer decides to re-optimize if the price change is greater than the threshold; otherwise not. We allow for heterogeneity among consumers. The re-optimizing threshold is thus individual-specific, denoted as L_k, where k is the index for consumers, from 1 to n. The price change is defined as $\Delta p_t = |p_t - p_{t-1}|$. Assume L_k is a random variable. For all $k = 1, 2, \cdots, n$, L_k is i.i.d. and follows distribution F. Given the setup of consumer's re-optimization process, we are able to obtain the probability of re-optimizing for each specific consumer k, which is $P(L_k < \Delta p_t) = F(\Delta p_t)$. Furthermore, the expected fraction of consumers who re-optimizes at time t can be expressed in terms of the above probability.

We then plug the above equation into the effective demand function and profit function.

$$\mathbb{E}\tilde{D}(p_t, p_{t-1}) = F(\Delta p_t)D(p_t) + (1 - F(\Delta p_t))D(p_{t-1}) \tag{3.5}$$

$$\pi(p_t|p_{t-1}) = (p_t - c_t)(F(\Delta p_t)D(p_t) + (1 - F(\Delta p_t))D(p_{t-1})) \tag{3.6}$$

We see that the last unspecified component in the profit function is the distribution F. Since F is the distribution for consumer's re-optimizing threshold, it should satisfy that $F(0) = 0$ (no negative threshold) and we also assume that F has a lower bound and an upper bound. A good candidate for F is uniform distribution. We will deploy uniform distribution in the following, i.e., $F \sim Uniform(0, U)$, where $U > 0$.

Now we are ready to derive the analytical solution for the retailer's profit maximization problem. We first write down the problem as

$$\max_{p_t} \quad (p_t - c_t)(F(\Delta p_t)D(p_t) + (1 - F(\Delta p_t))D(p_{t-1})) \tag{3.7}$$

Note that $F(\Delta p_t)$ is not differentiable at p_{t-1}, $p_{t-1} - U$ and $p_{t-1} + U$. The first-order condition for Equation (3.7) is given by

$$\begin{aligned} 0 &= F(\Delta p_t)D(p_t) + (1 - F(\Delta p_t))D(p_{t-1}) + \\ &(p_t - c_t)(\frac{dF(\Delta p_t)}{dp_t}D(p_t) + F(\Delta p_t)\frac{dD(p_t)}{dp_t} - \frac{dF(\Delta p_t)}{dp_t}D(p_{t-1})) \end{aligned} \tag{3.8}$$

where $p_t \notin \{p_{t-1}, p_{t-1} - U, p_{t-1} + U\}$. The following proposition states the analytical solution. To simplify the notation, denote $A \equiv -\beta_1\gamma(p_{t-1} - c_t)$, $B \equiv 3\beta_1\gamma(\beta_0 + \beta_1 p_{t-1})$ and $\gamma \equiv 1/U$.

Proposition

The profit maximizer p_t^* for Equation (3.7) is contained in the set of

$$\{p_{t-1} + \frac{A - \sqrt{A^2 - B}}{3\beta_1\gamma}, p_{t-1} - \frac{-A - \sqrt{A^2 + B}}{3\beta_1\gamma}, \frac{c_t - \beta_0/\beta_1}{2}, p_{t-1} - 1/\gamma, p_{t-1} + 1/\gamma\},$$

denoted as $\{p_t^1, p_t^2, p_t^3, p_{t-1} - 1/\gamma, p_{t-1} + 1/\gamma\}$.

Proof. See Appendix 1. □

By the above proposition, we know that Equation (3.7) has at most 5 profit maximizers. In the case when there are multiple profit maximizers (even though it is very unlikely), we assume the retailer chooses the smallest maximizer. By this assumption, we can construct a single-valued optimal pricing function $p_t^* = g(p_{t-1}, c_t; \beta_0, \beta_1, U)$. This optimal pricing function takes previous price and current cost as arguments and contains the parameters of the demand function and the distribution of consumer re-optimizing threshold.

3.2.3 An Example

To illustrate how our model works, we plug a set of parameters into the model to generate a simple example. For the simple demand function, we set $\beta_0 = 10$ and $\beta_1 = -2$, i.e., $D(p_t) = 10 - 2p_t$. The cost for the current period t is 1, i.e., $c_t = 1$. If there is no demand stickiness, it is easy to see that the optimal price is 3, i.e., $p_t^* = 3$ maximizes $D(p_t) \times (p_t - c_t)$. The maximized profit is 8. Now we consider how sticky demand affects retailer's optimal price and profit. We examine four cases where $U = 0.2, 0.15, 0.1$ and 0.05, respectively. The last period's price p_{t-1} takes values between 2.4 and 3.4. Figure 3.1 illustrates the response of p_t^* when p_{t-1} changes while other parameters remain the same. The x-axis is p_{t-1} and the y-axis is p_t^*. The solid line is a 45 degree line representing $p_t^* = p_{t-1}$, and the bubble line shows the value of p_t^*. We see that if p_{t-1} is far away from 3, the retailer will directly choose 3 as the optimal price. In such cases, p_t^* is far away from p_{t-1}, i.e., $\Delta p_t > U$, and there is no effect of demand stickiness. If p_{t-1} is close to 3, the retailer is able to leverage demand stickiness for a bigger profit. To be specific, the retailer tends to set a price that is a little bit higher than p_{t-1}. This results are consistent with [31]'s findings. The reason behind these results is that, with sticky demand, raising the price by a small amount leads to only a negligible loss of sales. Figure 3.2 shows how much more profit we can make with demand stickiness. We see that all four curves are above 8, which is the optimal profit that we can make without demand stickiness. If we compare the amount of the excess profit (relative to 8), we see that the higher the re-optimizing threshold is, the more excess profit the retailer makes.

3.3 Empirical Studies

In this section, we estimate the parameters in the theoretical model by using retail transactional data. The parameter of interest is the re-optimizing threshold distribution F, from which we learn how consumers react to price changes, or how sticky the demand is.

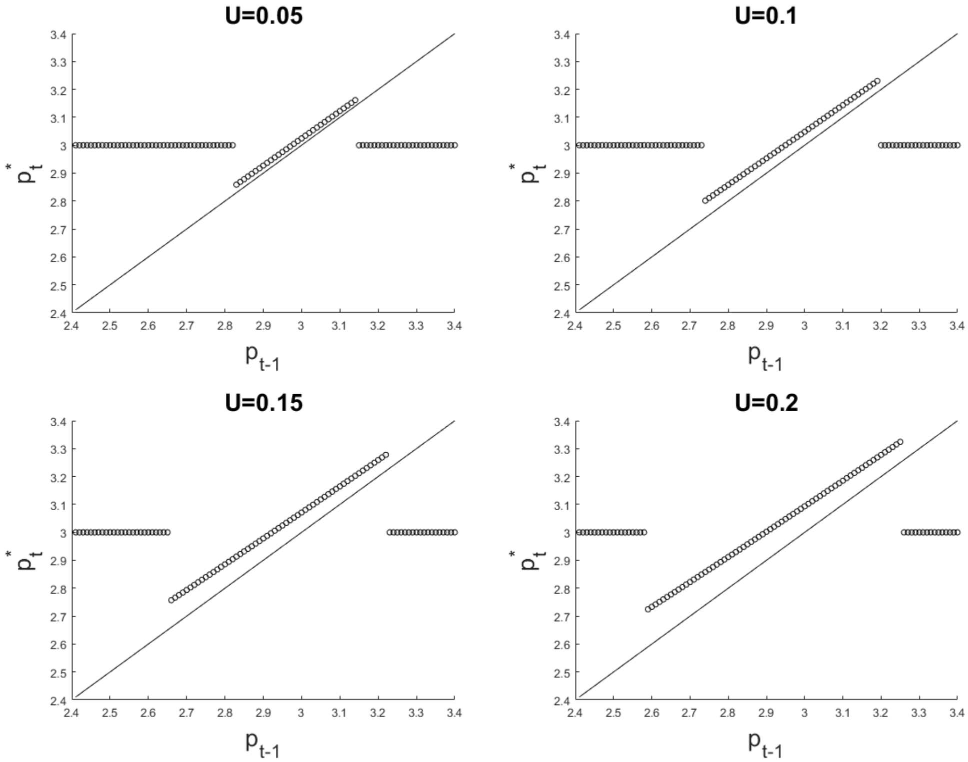

Figure 3.1: Effects of U on p_t^*

The data we use is Dominick's scanner data[4], which we briefly describe below.

3.3.1 Data Description

Dominick's was a large grocery chain in the greater Chicago area. The data set covers 400 weeks from 1989 to 1997, and the chain had about 100 stores. The data set contains weekly transactional data in each store for each product (identified by UPC) in 29 categories. A transaction data point includes the quantity of a certain product being sold during that specific week and the price at which the product is sold. We backtrack the average cost of each product in each week and each store using the price and profit data.

[4]See [51] for details about DominickâĂŹs data, which can be downloaded from http://research.chicagobooth.edu/marketing/databases/dominicks/index.aspx.

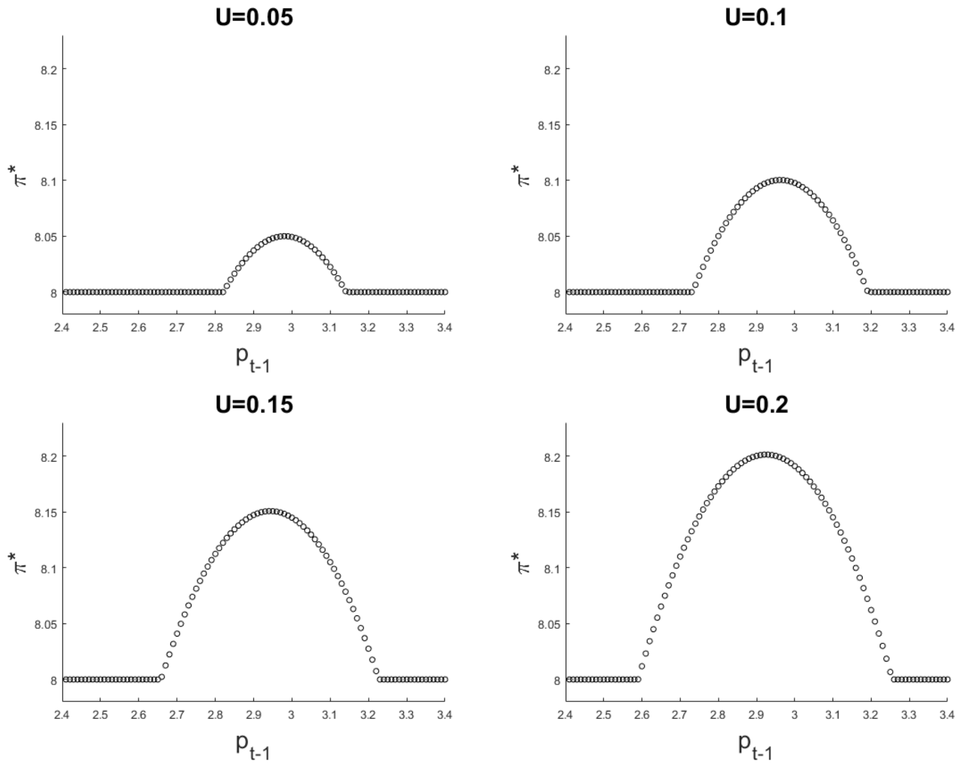

Figure 3.2: Effects of U on π^*

For each store, we can also observe demographics of the district in which the store is located. The demographics data includes basic social economics variables such as education and income, as well as some specific characteristics of shoppers such as the proportion of hurried shoppers[5].

Before we empirically estimate the model, we adjust the prices for inflation[6]. We exclude data from those weeks that fall into the holiday seasons (e.g., Thanksgiving, Christmas, New Year), as research has shown that Donimick's pricing behaviors are different during holidays [42]. We also exclude sales prices using Dominick's sales indicators[7].

[5]The proportion of hurried shopper refers to the percentage of consumers that go through the express lane in each store.

[6]We obtain the CPI for the Chicago area from http://www.bls.gov.

[7]DominickâĂŹs sales dummy is an imperfect measure of actual sales in that some sales are not recorded

With sale prices, retailers leverage all kinds of marketing tactics (e.g., store signage, color changes) to ensure that consumers react to the price reductions [53]. That is, with sale prices consumers will almost always re-optimize. Therefore, we use regular prices only to obtain a clean estimate of the re-optimizing threshold. In addition, changes in regular prices do not have predictable patterns, as factors that influence regular price changes (e.g., changes in wholesale prices, act of God) are less observable by consumers. Thus, the focus on regular prices is consistent with the assumption of myopic consumers in our theoretical model.

3.3.2 Estimation Method

In this section, we discuss the method for estimating the parameters in the theoretical model. From the data description, we know that the data structure is of a panel structure. We have multiple products, indexed as $i = 1, \cdots, N$, and multiple time periods, indexed as $t = 1, \cdots, T$. The observables are sales ($q_{i,t}$), price ($p_{i,t}$) and cost ($c_{i,t}$).

3.3.2.1 Estimation of the Simple Demand Functions

First, we estimate the parameters in the simple demand functions. Note that for different products, their simple demand functions are not necessarily identical. Thus, the demand function estimation is product-specific. The parameters to be estimated are β_0 and β_1 for each product i, which are denoted as $\beta_{i,0}$ and $\beta_{i,1}$ respectively for exposition purposes. From Equation (3.5), we have that

$$E\tilde{D}(p_{i,t}, p_{i,t-1}) = F(\Delta p_{i,t})D(p_{i,t}) + (1 - F(\Delta p_{i,t}))D(p_{i,t-1})$$

which are β_0 and β_1 for product i. Note that the unknown value of the function F will lead to non-identification of $\beta_{i,0}$ and $\beta_{i,1}$. To get rid of this problem, we use sub-sample where

[52]. Therefore, our estimation of demand stickiness is a conservative one.

$p_{i,t} = p_{i,t-1}$. When $p_{i,t} = p_{i,t-1}$, we have that $\Delta p_{i,t} = 0$ and $F(\Delta p_{i,t}) = 0$. The equation above reduces to

$$E\tilde{D}(p_{i,t}, p_{i,t-1}) = D(p_{i,t-1}) = D(p_{i,t})$$

We assume $E(q_{i,t}) = E\tilde{D}(p_{i,t}, p_{i,t-1}) = D(p_{i,t})$. By adding a structural error term $u_{i,t}$, we write the econometrics model as

$$q_{i,t} = D(p_{i,t}) + u_{i,t} = \beta_{i,0} + \beta_{i,1} p_{i,t} + u_{i,t} \tag{3.9}$$

where $u_{i,t} \sim i.i.d.N(0, \sigma^2_{i,u})$. For each product i, we are now able to estimate $\beta_{i,0}$ and $\beta_{i,1}$. The estimation results in a set of $\{\beta_{i,0}\}_{i=1}^N$ and a set of $\{\beta_{i,1}\}_{i=1}^N$, where N is the number of products.

3.3.2.2 Estimation of Re-Optimizing Thresholds

As described in Section 3.2, the re-optimizing threshold only plays a role in the second stage of the retailer's decision process. To identify the re-optimizing threshold, we focus on the cases in which the retailer enters into the second stage, i.e., the retailer chooses to change the price. Thus, we use another sub-sample where $p_{i,t} \neq p_{i,t-1}$ for the estimation.

Recall that the re-optimizing threshold L_k is a random variable, where k is the index for consumers. A good way to present L_k is to report its mean, denoted as $\bar{L}$. For each product i, we expect a product-specific $\bar{L}_i$. However, reporting product-specific $\bar{L}_i$ for all products will be too cumbersome, without providing much insights. Our main interest is the re-optimizing threshold at the category level and store level. To pool across products i on category level and store level, we introduce relative re-optimizing threshold, denoted as $\bar{L}_R$, assuming that the relative re-optimizing threshold is identical across products within the same category and same store. The relative re-optimizing threshold is the threshold adjusted by the specific product's average price, i.e., $\bar{L}_R = \frac{\bar{L}_i}{\bar{p}_i}$, where $\bar{p}_i$ is the average price. The average price is obtained by $\bar{p}_i = \frac{1}{T}\sum_{t=1}^{T} p_{i,t}$. Henceforth, the product index i is within a certain category and a certain store, and the estimation of $\bar{L}_R$ is conditional on category and store. From Section 3.2, we know that the theoretical optimal price is given by

$$p^*_{i,t} = g(p_{i,t-1}, c_{i,t}; \beta_{i,0}, \beta_{i,1}, U_i)$$

Recall that L_k is assumed to be uniformly distributed and U_i is the upper bound. We immediately have $\bar{L}_i = \frac{U_i}{2}$ and $\bar{L}_R = \frac{U_i}{2\bar{p}_i}$. We express the above equation in terms of the relative re-optimizing threshold.

$$p^*_{i,t} = g(p_{i,t-1}, c_{i,t}; \beta_{i,0}, \beta_{i,1}, 2\bar{p}_i\bar{L}_R)$$

We assume that $E(p_{i,t}) = p^*_{i,t}$. By adding a structural error term $\epsilon_{i,t}$, we write the econometrics model.

$$p_{i,t} = p^*_{i,t} + \epsilon_{i,t} = g(p_{i,t-1}, c_{i,t}; \beta_{i,0}, \beta_{i,1}, 2\bar{p}_i\bar{L}_R) + \epsilon_{i,t}$$

where $\epsilon_{i,t} \sim i.i.d. N(0, \sigma^2_\epsilon)$. Plug in the estimates of $\beta_{i,0}$ and $\beta_{i,1}$ from the previous section, we have

$$p_{i,t} = g(p_{i,t-1}, c_{i,t}; \hat{\beta}_{i,0}, \hat{\beta}_{i,1}, 2\bar{p}_i\bar{L}_R) + \epsilon_{i,t} \tag{3.10}$$

Note that the only unknown parameter is $\bar{L}_R$. Since the function g is nonlinear in $\bar{L}_R$, we use the method of nonlinear least squares to estimate $\bar{L}_R$.

To check the robustness of our results, we alternatively estimate the re-optimizing threshold as it is directly revealed in consumer's purchase behavior. From Section 3.2, we have that

$$E\tilde{D}(p_{i,t}, p_{i,t-1}) = F(\Delta p_{i,t})D(p_{i,t}) + (1 - F(\Delta p_{i,t}))D(p_{i,t-1})$$

Writing the above equation in terms of parameters, we have

$$E\tilde{D}(p_{i,t}, p_{i,t-1}) = \beta_{i,0} + \beta_{i,1}(F(\Delta p_{i,t}; U_i)p_{i,t} + (1 - F(\Delta p_{i,t}; U_i)p_{i,t-1})$$

Plugging in the relationship between U_i and $\bar{L}_R$, we have

$$E\tilde{D}(p_{i,t}, p_{i,t-1}) = \beta_{i,0} + \beta_{i,1}(F(\Delta p_{i,t}; 2\bar{p}_i\bar{L}_R)p_{i,t} + (1 - F(\Delta p_{i,t}; 2\bar{p}_i\bar{L}_R)p_{i,t-1})$$

Assuming $E(q_{i,t}) = E\tilde{D}(p_{i,t}, p_{i,t-1})$ and adding a structural error term $v_{i,t}$, we write the econometrics model as

$$q_{i,t} = \beta_{i,0} + \beta_{i,1}(F(\Delta p_{i,t}; 2\bar{p}_i\bar{L}_R)p_{i,t} + (1 - F(\Delta p_{i,t}; 2\bar{p}_i\bar{L}_R)p_{i,t-1}) + v_{i,t}$$

where $v_{i,t} \sim i.i.d.N(0, \sigma^2_{i,v})$. Plugging in the estimates of $\beta_{i,0}$ and $\beta_{i,1}$, we have

$$q_{i,t} = \hat{\beta}_{i,0} + \hat{\beta}_{i,1}(F(\Delta p_{i,t}; 2\bar{p}_i\bar{L}_R)p_{i,t} + (1 - F(\Delta p_{i,t}; 2\bar{p}_i\bar{L}_R)p_{i,t-1}) + v_{i,t} \qquad (3.11)$$

Note that the only unknown parameter is $\bar{L}_R$. Since the function F is nonlinear in $\bar{L}_R$, we use the method of nonlinear least squares to estimate γ.

3.3.3 Results

Recall that each relative re-optimizing threshold $\bar{L}_R$ is estimated for a certain category and a certain store. The first way to report these estimates is by summarizing the thresholds across stores, i.e., to report distributional characteristics (mean and standard deviation). Table 3.1 reports the estimation results for 12 categories. These categories consist of six that represent relatively less consumable products and six that represent relatively more consumable products. We expect that the products that are less consumable can be stockpiled, and that consumers who bought these products at regular prices are those that face higher re-optimization costs. Therefore, comparing the re-optimizing threshold between the two groups (with six categories each) allows us to validate our logic for demand stickiness that is due to consumer re-optimizing cost. Specifically, we expect that the re-optimizing threshold should on average be bigger for the more consumable than the less consumable products. As each category contains a large number of stores, we summarize the results as mean and stand deviation across stores. The first two columns in Table 3.1 are the results from the main analysis and the last two columns are those from the alternative estimation approach. The two approaches give similar results, providing evidence for the robustness of our structural model.

On average, the relative re-optimizing threshold ranges between 0.01 and 0.29. The majority, i.e., 19/24 or 79%, of the relative re-optimizing thresholds are at or above 0.05. As expected, the threshold is indeed bigger for the six less consumable categories than the six more consumable categories, confirming our logic based on consumer re-optimizing cost.

Table 3.1: Estimation Results for 12 Categories

	Main		Alternative	
	Mean	Std	Mean	Std
Bath Soap	0.18	0.13	0.19	0.38
Dish Detergent	0.15	0.08	0.05	0.22
Grooming Products	0.17	0.06	0.10	0.11
Shampoos	0.18	0.07	0.16	0.08
Soap	0.11	0.07	0.05	0.10
Toothbrushes	0.29	0.16	0.10	0.11
Cookies	0.05	0.03	0.05	0.04
Crackers	0.06	0.06	0.06	0.08
Canned Soup	0.08	0.04	0.01	0.01
Front-End-Candies	0.03	0.05	0.05	0.05
Soft Drinks	0.03	0.04	0.02	0.02
Snack Crackers	0.07	0.05	0.04	0.06

Table 3.2 shows the relationship between relative re-optimizing threshold and demographics. The demographic variables we pick here are % of college graduates and above (educ), income (income), % of whites (ethnic), average household size (hsizeavg), population density (density), % of hurried shopper (shphurr), and % of non-working women with children (nwrkch). We include these variables in our analysis with the expectation that they are correlated with the consumer's re-optimizing cost. We estimate their effects with a linear regression with category fixed effects. We log-transform the threshold as our dependent variable, to capture the scaling (vs. shifting) effects of consumer demographics. We see that the threshold is positively correlated with the proportion of hurried shopper, and negatively correlated with the average household size and the proportion of college graduates and above. These results are intuitively appealing and further validate our logic based on consumer's re-optimizing cost. Specifically, hurried shoppers do not have enough time to remember or compare a price, which leads to a high level of re-optimizing cost, resulting in a higher threshold. For people in a large household with relative tight budge, a small price change in grocery items can be more consequential. Therefore, these consumers experience higher re-optimizing benefits, resulting in a lower threshold. Finally, people with more education should have more cognitive resources to process price information than the rest of the population. Those consumers experience lower cost of remembering and comparing prices, leading to lower re-optimizing threshold.

We also estimate the effects of store and product characteristics on the re-optimizing threshold. Dominick's groups its stores into four tiers: Cub Fighters, Low, Medium, High. We estimate the threshold for each tier. Table 3.3 shows the effects of price tiers on the threshold. We also report the average income for each price tier. The average income is the highest in the Cub Fighter tier, while the tier with the second highest income is medium followed by low and high.[8] We see that in the case of Cub Fighters, the threshold is the

[8]This hints at the possibility that Dominick's was likely to open a store next to a Cub Foods to fight for

Table 3.2: Regression on Demographics

	$log(\bar{L}_R)$
educ	-2.881
	(-1.89)*
income	-0.000
	(-0.02)
ethnic	0.968
	(1.54)
hsizeavg	-2.098**
	(-2.54)
density	-17.23
	(-0.18)
shphurr	12.80**
	(2.17)
nwrkch	-0.474
	(-0.16)
N	995

smallest even though the income is the highest, testifying to the power of competition on reducing the re-optimizing costs. The threshold for the Low price tier is similarly low, which is not surprising given that this tier has the lowest income.

Finally, Table 3.4 shows the re-optimizing threshold comparing national brands and private labels. We follow [42] and [54] and use pairs of national brands and private labels that match on sizes and quality. We see that there is no significant difference between national brands and private labels.

3.4 Simulation

In this section, we present an application of demand stickiness in a dynamic pricing setting. Traditional studies of dynamic pricing assume that consumers will respond to any price change, no matter how small it is. If we relax this assumption by introducing sticky demand, the traditional dynamic pricing strategy may no longer be the optimal

high income consumers and confirms the speculation that retailers may take the advantage of constraints faced by low income consumers in terms of storage, transportation, and cognitive resources.

Table 3.3: Effects of Price Tiers

	$\bar{L}$	Income ($\times 1000$)
Cub Fighter	0.111	46.13
Low	0.123	42.04
Med	0.120	44.01
High	0.112	38.74

Table 3.4: National Brand versus Private Label

	National Brand	Private Label
Bottled Juice	0.13	0.11
Cereals	0.15	0.05
Cheese	0.11	0.11
Cookies	0.09	0.11
Crackers	0.09	0.15
Canned Soup	0.09	0.14
Dish Detergent	0.13	0.10
Frozen Entrees	0.12	0.10
Frozen Juices	0.18	0.15
Fabric Softener	0.11	0.08
Grooming Products	0.10	0.21
Laundry Detergents	0.20	0.13
Oatmeal	0.11	0.11
Snack Crackers	0.06	0.08
Canned Tuna	0.10	0.11

pricing strategy. Therefore, we propose a new dynamic pricing strategy to accommodate sticky demand. Several simulation cases are set up to examine the performance (in terms of distributional characteristics of revenue) of both traditional and new dynamic pricing strategies. In addition, we also consider the performance of fixed pricing as a bench mark.

3.4.1 Simulation Setup

Suppose there are 100 units of a perishable good in the inventory. A retailer has T weeks to sell them. After T weeks, the salvage value is 0. T takes three values: 10, 15 and 20. The expected effective demand is defined in the same way as in Section 3.2, i.e., $E\tilde{D}(p_t, p_{t-1}) = F(\Delta p_t)D(p_t) + (1 - F(\Delta p_t))D(p_{t-1})$. Assume that the effective demand follows a Poisson distribution with mean as $E\tilde{D}(p_t, p_{t-1})$, i.e., $\tilde{D}_t(p_t, p_{t-1}) \sim Poisson(E\tilde{D}(p_t, p_{t-1}))$. For the parameters in the simple demand function, we use the estimates from the previous section. Since every product in every store has a unique simple demand function, we randomly pick a product in the category of crackers and a store. Set $\beta_0 = 16.67$ and $\beta_1 = -4.63$, i.e., $D(p_t) = 16.67 - 4.63p_t$. The mean re-optimizing threshold $\bar{L}_R$ takes three values: 0.1, 0.25 and 0.4. Assume that the discount factor is 1.

For fixed pricing, there is only one parameter to be determined, which is the fixed price p. The revenue maximization problem is given by

$$\max_{p} \quad \sum_{t=1}^{T} p \times (Pr(I_t)E(\tilde{D}_t(p,p)|I_t) + (1 - Pr(I_t))E(max\{100 - \sum_{s=1}^{t-1} \tilde{D}_s(p,p)), 0\}) \tag{3.12}$$

where condition I_t is $\sum_{s=1}^{t} \tilde{D}_s(p,p) \leq 100$. It means that the realization of demand in each period cannot exceed the remaining inventory. If there are too many customers arriving, the product will be out of stock, in which case we set the demand as the remaining inventory available. The optimal price for fixed pricing is numerically solved with precision of one decimal.

We use dynamic programming to solve for the optimal prices in the cases when dynamic pricing strategy is adopted. Denote $V^{A}_{t,G_t}(p_t)$ as the value function for the case with a re-optimizing threshold of zero, where t is the current time period, G_t is the remaining inventory at time t and p_t is the price to be set in this time period. The state variables are t and G_t, and the control variable is p_t. Denote $V^{I}_{t,G_t,p_{t-1}}(p_t)$ as the value function for the case with a non-zero re-optimizing threshold, where t is the current time period, G_t is the remaining inventory at time t, p_{t-1} is the last period's price and p_t is the price to be set in this time period. The state variables are t, G_t and p_{t-1}, and the control variable is p_t. When $t = 1$, we set $p_0 = \infty$, so there is no sticky demand effect in the first time period.

We adopt the algorithm of backward induction. The Bellman equation for the case with a non-sticky demand is given by

$$V^A_{t,G_t} = \max_{p_t} \quad E(p_t \tilde{D}_t(p_t, p_t) + V^A_{t+1,G_{t+1}} | \tilde{D}_t(p_t, p_t) \leq G_t) \tag{3.13}$$

After T weeks, the good has no value, i.e.

$$V^A_{T+1,G_{T+1}} = 0, \ \ \forall G_{T+1} \tag{3.14}$$

Note that the effective demand function $\tilde{D}_t(p_t, p_t)$ does not depend on last period's price under the assumption that demand is non-sticky. On the other hand, the Bellman equation for the case with a sticky demand is given by

$$V^I_{t,G_t,p_{t-1}} = \max_{p_t} \quad E(p_t \tilde{D}_t(p_t, p_{t-1}) + V^I_{t+1,G_{t+1},p_t} | \tilde{D}_t(p_t, p_{t-1}) \leq G_t) \tag{3.15}$$

After T weeks, the good has no value, i.e.,

$$V^I_{T+1,G_{T+1},p_T} = 0, \ \ \forall G_{T+1}, p_T \tag{3.16}$$

These dynamic programming problems are solved numerically. The optimal price has a precision of one decimal.

3.4.2 Simulation Results

Tables 3.5, 3.6 and 3.7 report the revenues generated from 1,000 simulations with $T = 10$, 15 and 20. The revenues are summarized in mean (over the bracket), standard deviation (in round bracket) and frequency of generating the highest revenue (in square bracket). The mean of the revenues is a simple measure of a pricing strategy's performance. Under risk-aversion, the standard deviation also play an important role in determining the performance. The frequency of generating the highest revenue is a direct indicator of the best pricing strategy. In each case, we report the results from the three pricing strategies: fixed pricing, dynamic pricing with non-sticky demand, and dynamic pricing with sticky demand, and from three values of re-optimizing threshold: $\bar{L} = 0.1$, 0.25 and 0.4.

When $T = 10$, dynamic pricing with sticky demand has the highest mean and frequency of generating the highest revenue, while fixed pricing has the lowest standard errors. Fixed pricing and dynamic pricing with non-sticky demand perform similarly on all three aspects. As the time horizon T becomes longer, for all pricing strategies, the means increase and standard deviations decrease, i.e., all of the pricing strategies perform better. The advantage of dynamic pricing against fixed pricing becomes larger as T increases. If we look at the variation of the re-optimizing threshold $\bar{L}$, the advantage of dynamic pricing with sticky demand against dynamic pricing with non-sticky demand becomes larger as $\bar{L}$ increases.

Table 3.5: Simulated Revenues with $T = 10$

	$\bar{L} = 0.1$	$\bar{L} = 0.25$	$\bar{L} = 0.4$
Fixed	149.56	149.82	149.63
	(16.00)	(15.69)	(15.61)
	[0.30]	[0.28]	[0.25]
Dynamic non-sticky	150.30	149.82	149.58
	(16.24)	(16.38)	(16.47)
	[0.31]	[0.28]	[0.27]
Dynamic sticky	152.61	155.46	157.55
	(16.46)	(16.60)	(16.90)
	[0.38]	[0.43]	[0.48]

Figure 3.3 shows the price trajectory in one single simulation with $T = 20$ and $\bar{L} = 0.25$. The fixed pricing sets $p = 2.5$ as the optimal price. Compared to the fixed pricing,

Table 3.6: Simulated Revenues with $T = 15$

	$\bar{L} = 0.1$	$\bar{L} = 0.25$	$\bar{L} = 0.4$
Fixed	208.07	208.04	208.37
	(14.29)	(14.32)	(14.44)
	[0.24]	[0.20]	[0.19]
Dynamic non-sticky	210.76	210.31	210.77
	(14.75)	(15.34)	(15.25)
	[0.33]	[0.29]	[0.28]
Dynamic sticky	213.83	216.84	218.51
	(14.02)	(14.40)	(14.03)
	[0.43]	[0.50]	[0.53]

Table 3.7: Simulated Revenues with $T = 20$

	$\bar{L} = 0.1$	$\bar{L} = 0.25$	$\bar{L} = 0.4$
Fixed	242.11	242.44	242.15
	(12.71)	(12.48)	(12.83)
	[0.17]	[0.15]	[0.13]
Dynamic non-sticky	246.78	246.30	246.24
	(11.70)	(12.23)	(12.26)
	[0.34]	[0.30]	[0.29]
Dynamic sticky	249.56	251.88	253.12
	(11.01)	(11.15)	(10.93)
	[0.49]	[0.54]	[0.59]

the dynamic pricing with non-sticky demand sets higher prices in early periods and lower prices in late periods. For dynamic pricing with sticky demand, if we smooth out the regular cycles and focus on the long-term trend, we see a similar pattern as that shown for traditional dynamic pricing (higher in early periods and lower in late periods). When we examine those cycle, we see that when the price goes up, it takes two to three periods for the price to increase, and when the price goes down, it takes only one period for the price to decrease. That is, price increases are slow and price decreases are fast. These results are consistent with our theoretical model implication and empirical findings in the literature

[31, 52].

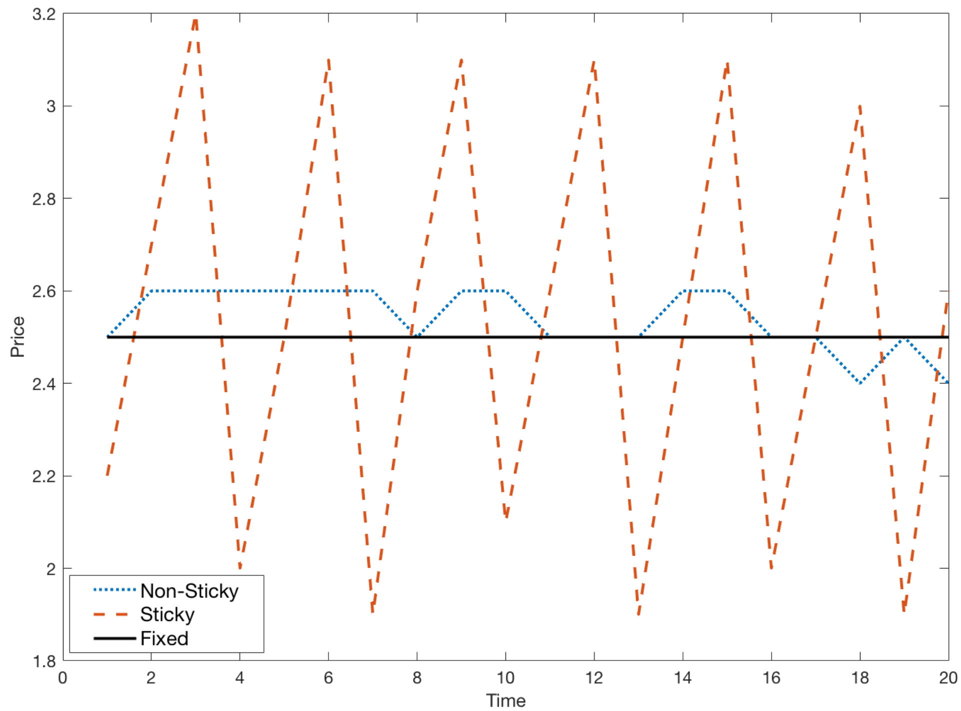

Figure 3.3: Price Trajectory

4. NONPARAMETRIC CONDITIONAL QUANTILE FUNCTION ESTIMATION FOR TIME SERIES DATA

Compared with a conditional mean function, a conditional quantile function can provide a comprehensive picture about the conditional distribution of a dependent variable conditional on explanatory variables. Hence, the study of conditional quantile function has its empirical appeals. For this reason there is a rich literature in studying estimation of conditional quantile functions. For independent and identically distributed (i.i.d.) data, the study of conditional quantile function estimation stems from the seminal work of [55] who study conditional quantile function in a parametric model framework. More recently, a growing literature thrive to study the conditional quantile function in a nonparametric framework, requested by the empirical needs of practitioners, such as [56], [57], [58],[59], [60], [61], [62], [63], [64], and [65].

Many existing nonparametric conditional quantile estimators are indirect estimators since they are solutions to equations that indirectly deliver a quantile. Among literature on indirect nonparametric conditional quantile estimators, [56], [57], [58], [59], [60], and [63] use the check function to indirectly estimate conditional a quantile function non-parametrically. [61] indirectly estimate conditional quantile function nonparametrically by inverting an estimated conditional CDF. In addition to indirect estimation methods in conditional quantile estimation, [65] initiates a direct nonparametric conditional quantile estimation method, which has two advantages over indirect methods. First, the pointwise bias of the direct estimator contains one less term than the indirect estimator, hence the direct estimator has smaller minmax bias risk. Second, when the explanatory variable data support is unbounded, the variance of the direct estimator is smaller in magnitude than the check function based counterpart in the tail regions. Hence, the direct estimator may offer substantial finite sample efficiency gains relative to the indirect estimator in the tail regions.

The distributions of many financial variables are known to have fat tails. Hence, the direct nonparametric conditional quantile estimator may outperform the indirect estimator in tail regions. However, if one wants to apply direct nonparametric conditional quantile estimation method to forecast some financial variables, one should establish the validity of the direct nonparametric conditional quantile estimator for dependent data. In particular, one needs to study the asymptotic distribution of the conditional quantile estimator. Many authors have studied the theoretical properties of nonparametric estimators for dependent data by establishing Central Limit Theorems (CLT) for various type dependent data. For example, [66] establishes CLTs for degenerate U-statistics of m-dependent processes. [67], [68] and [69] consider degenerate U-statistics of absolutely regular processes and other weakly dependent processes. [70] prove the asymptotic normality of the integrated squared error of the kernel density estimator for absolutely regular processes. [71] provide a general CLTs for second order degenerate U-statistics with variable kernels for absolutely regular processes. Using Central Limit Theorems for dependent observations, [72] study the nonparametric regression model $Y_t = \theta(Z_t) + X_t$, where $\theta(.)$ is an unknown function and $\{X_t\}$ follows an autoregressive process. Under a set of high level assumptions [73] establishes the $\sqrt{n}$-consistency and asymptotic normality of MINPIN estimators, which are estimators that minimize a criterion function that may depend on a preliminary infinite dimensional nuisance parameter estimator. [74] extends the $\sqrt{n}$-consistency and asymptotic normality results of a partially linear model of [75] for independent observations to weakly dependent processes.

In this paper we study the theoretical properties of the direct nonparametric conditional quantile estimator proposed by [65] with weakly dependent observations. We prove the asymptotic normality of the direct nonparametric conditional quantile estimator under absolutely regular processes. We derive the asymptotic distribution of our proposed estimator and compare the finite sample performance of our estimator and traditional check function estimator by Monte Carlo simulations. We show that our proposed direct method outperforms conventional check function based method in terms of out-of-sample predictions. We then apply the direct nonparametric estimation method to forecast U.S. housing return based on S&P/Case-Shiller House Price Indices and compare the forecasting results with that of some indirect estimators. The comparison results show that the forecasting values of housing return by direct estimation method indeed outperforms the indirect counterpart, especially for data in the tail regions.

The paper is organized as follows: we study the asymptotic normality of the direct nonparametric conditional quantile estimator under absolutely regular processes in Secion 2. Section 3 conducts the Monte Carlo simulations to study the finite sample performance of our proposed direct method. In Section 4, we apply the direct estimation method to forecast U.S. monthly housing returns and compare the forecast result with its check function based counterpart. The proofs of the main results are presented in the Appendix 2.

4.1 Direct Nonparametric Conditional Quantile Estimator

We consider the following location-scale model:

$$Y_t = a(X_t) + b(X_t)\,\varepsilon_t, \quad t = 1, \dots.., n$$

where $a(X_t)$ and $b(X_t) \geq 0$ are unknown smooth location and scale functions, the idiosyncratic error ε_t is i.i.d with zero mean and a unit variance with an unknown distribution function $F_\varepsilon(\cdot)$. We consider the problem of estimating the τ-th conditional quantile function with $\tau \in [\underline{\tau}, \bar{\tau}] \equiv \Lambda$, where $0 < \underline{\tau} < \bar{\tau} < 1$.

4.1.1 The Local Constant Estimator

Following [65], we estimate the conditional τ-th quantile function of Y_t given X_t evaluated at $X_t = x$ by

$$\hat{q}_\tau(x) = \frac{\frac{1}{nh}\sum_{t=1}^{n} Q_{N(Y_t,\hat{b}(X_t)),\hat{\delta}} K((X_t - x)/h)}{\hat{f}(x)}, \tag{4.1}$$

where $Q_{N(\mu,\sigma),\delta} = \mu + \sigma Q_{N(0,1),\delta}$ denotes the $\delta \in (0,1)$ quantile of a Gaussian random variable with mean μ and standard deviation σ, $Q_{N(0,1),\delta}$ is the $\delta \in (0,1)$ quantile of a standard Gaussian random variable, $\hat{\delta} \in (0,1)$ is an estimated parameter that is determined by the least squares cross validation data-driven method that we will describe below, $\hat{f}(x) = (nh)^{-1}\sum_{t=1}^{T} K((X_t - x)/h)$ is the kernel estimator of the marginal density for X_t evaluated at x, $\hat{b}(X_t) = \sqrt{\hat{b}^2(X_t)}$ is a kernel estimator of the conditional standard deviation of Y_t, $\hat{b}^2(X_t) = \hat{E}\{[Y_t - \hat{E}(Y_t|X_t)]^2|X_t\}$ is the kernel estimator of the conditional variance of Y_t given X_t. The reason that $\hat{q}_\tau(x)$ defined in 4.1 is a consistent estimator of $q_\tau(x)$ is that for any $\tau \in (0,1)$, there exist a $\delta_0 = \delta_0(\tau)$ such that $Q_{N(0,1),\delta_0} = Q_{\epsilon,\tau}$, where $Q_{\epsilon,\tau}$ is the τ^{th} quantile of ϵ. As in [65], one can show that if $\hat{\delta} = \delta_0 + Op(1)$, then $\hat{q}_\tau(x) = q_\tau(x) + Op(1)$. Indeed, we will show that $\hat{\delta} = \delta_0(\tau) + Op(1)$ in this paper.

This nonparametric conditional quantile local constant estimator requires two steps. In the first step, we estimate conditional moments of Y_t using a second order kernel function: $\hat{E}(Y_t^l|X_t) = (nh_1)^{-1}\sum_{s=1}^{n} Y_s^l \tilde{K}_{st}/\tilde{f}(X_t)$ for $l = 1, 2$, where $\tilde{f}(X_t) = (nh_1)^{-1}\sum_{s=1}^{n} \tilde{K}_{st}$, $\tilde{K}_{st} = K((X_s - X_t)/h_1)$. This gives us kernel estimators of $\hat{a}(X_t)$ and $\hat{b}(X_t)$. In the second step, we select η and h simultaneously by minimizing the following objective function:

$$(\hat{\eta}, \hat{h}) = \arg\min_{(\eta,h)\in[\underline{\eta},\overline{\eta}]\times H_n} \frac{1}{n}\sum_{t=1}^{n} \rho_\tau(Y_t - \hat{q}_{\tau,-t}(X_t))M(X_t) \tag{4.2}$$

where $\rho_\tau(.)$ is the check function, i.e., $\rho_\tau(z) = z[\tau - 1(z \le 0)]$, $\eta \overset{def}{=} Q_{N(0,1),\delta} \equiv$

$\sqrt{2}erf^{-1}(2\delta - 1)$,[1] $[\underline{\eta}, \overline{\eta}]$ is the support of η, which corresponds to $[\underline{\tau}, \overline{\tau}]$, the support of τ. Here, as in [65], we reparametrized the parameter from δ to $\eta = \eta(\delta) = \sqrt{2}erf^{-1}(2\delta - 1)$ to simplify notation.

Let $H_n = [c_1 n^{-1/5}, c_2 n^{-1/5}]$ be the support of h with $0 < c_1 < c_2 < \infty$, i.e., $h \in H_n$, let $M(X_t)$ be a trimming function that trims out data near the boundary, and $\hat{q}_{\tau,-t}(X_t)$ is the leave-one-out τ-th conditonal quantile estimator:

$$\hat{q}_{\tau,-t}(X_t) = \frac{\frac{1}{nh}\sum_{s\neq t}^{n} Q_{N(Y_s,\hat{b}(X_s)),\delta} K_{st}}{\hat{f}_{-t}(X_t)}$$

where $K_{st} = K((X_s - X_t)/h)$, $\hat{f}_{-t}(X_t) = (nh)^{-1}\sum_{s\neq t}^{T} K((X_s - X_t)/h)$ is the leave-one-out kernel estimator of $f(X_t)$.

We use a two-step estimation procedure to estimate the τ-th nonparametric conditional quantile as in [65]. Also as in [65], we choose undersmoothing in the first step, i.e., $h_1 = o(h)$ and $nh_1 \to \infty$ as $n \to \infty$, so that asymptotically the first step estimation error has negligible effects on the second step estimation, and the asymptotic distribution of $\hat{q}_\tau(x)$ remains the same if we replace $\hat{b}(X_t)$ with $b(X_t)$.

We assume that data $\{(X_t, Y_t)\}_{t=1}^n$ follows a weakly dependent, absolutely regular (β-mixing) processes. We give a formal definition of an absolutely regular processes now. we say that $\{\mathcal{W}_t\} \equiv \{Y_t, X_t\}_{t=1}^n$ follows a strictly stationary, absolutely regular process if as $r \to \infty$,

$$\beta_r = \sup_{s\in N} E\left[\sup_{\mathcal{A}\in\mathcal{M}_{s+r}^{\infty}(\mathcal{W})} \left\{|P(\mathcal{A}|\mathcal{M}_{-\infty}^{s}(\mathcal{W})) - P(\mathcal{A})|\right\}\right] \to 0,$$

where $\mathcal{M}_s^t(\mathcal{W})$ denotes $\sigma(\mathcal{W}_s, \dots, \mathcal{W}_t)$, the sigma algebra generated by $(\mathcal{W}_s, \dots, \mathcal{W}_t)$, for $s \leq t$.

To derive the asymptotic distribution of $\hat{q}_\tau(x)$, we make the following assumptions:

[1] Here, $erf^{-1}(\cdot)$ is the inverse error function and the error function is defined as $erf(x) = \int_0^x e^{-v^2} dv$.

A.1 $\{Y_t, X_t\}$, $t = 1, 2, \ldots, n$ is a strictly stationary absolutly regular process (β-mixing) with the β-mixing coefficient β_r, satisfying $\beta_r^{\theta/(2+\theta)} = O(r^{-2+\epsilon})$, for some $0 < \epsilon < 1, 0 < \theta \leq 1/2$.

A.2 The idiosyncratic error $\{\varepsilon_t\}_{i=1}^n$ i.i.d. with zero mean and unit variance.

A.3 Let $\mathcal{S}$ be the compact support of the trimming function $M(X_t)$, i.e. $M(X_t) \neq 0$ if and only if $X_t \in \mathcal{S}$, $\inf_{x \in \mathcal{S}} f_X(x) > 0$, $\inf_{x \in \mathcal{S}} b(x) > 0$.

A.4 $q_\tau(.) \in \mathcal{G}_3$, $f(.) \in \mathcal{G}_2$, $a(.) \in \mathcal{G}_3$, $b(.) \in \mathcal{G}_5$, where $\mathcal{G}_l$ denotes the class of functions that are l-times continuously differentiable at $x \in \mathcal{S}$.

A.5 The kernel function $K(.)$ is a bounded, symmetric density function, with

$$\int K(v) v^4 dv < \infty.$$

The smoothing parameters satisfying: $h_1 = O(n^{-1/5})$ and $h = O_e(n^{-1/5})$, so that $nh_1 \to \infty$, $nh \to \infty$, $nh_1^5 \to 0$, $nh^6 \to 0$ as $n \to \infty$.

Assumption A.1 assumes that the $\{Y_t, X_t\}$ process is absolutely regular (β mixing) process with some restrictions on the decay rate of its mixing coefficient β_r. This is not a stringent assumption as discussed in [74]. Assumption A.2 imposes the i.i.d difference structure on $\{\varepsilon_t\}$. Assumption A.3 gives some smoothness and moment conditions, similar to those given in [65]. Assumption A.4 requires that the density function $f_X(\cdot)$ and the scale function $b(\cdot)$ are both bounded on the trimming set. Assumption A.5 ensures that the bias of the kernel estimates is of order h^2 and imposes some restrictions on the smooth parameter h_1 and h.

The next two propositions give the asymptotic results for the cross validation selected parameters $\hat{\eta}$ and $\hat{h}$ defined in (4.2).

Proposition 1. *Define* $\eta_0 = \sqrt{2} erf^{-1}(2\delta_0 - 1) = Q_{N(0,1),\delta_0}$*, where* $\delta_0 = \delta(\tau_0)$ *such that* $Q_{N(0,1),\delta_0} = Q_{\epsilon,\tau_0}$*. Let* $(\hat{\eta}, \hat{h})$ *be the value of* (η, h) *that minimizes the objective function*

$CV(\eta, h)$, then under assumptions A.1-A.5, we have

$$\hat{\eta} - \eta_0 = -B_{11}h^2 + O_p(n^{-\frac{1}{2}})$$

where $B_{11} = E\left[M_t f_{e,t} B_{LC,t} b_t\right] / E[M_t f_{e,t} b_t^2]$, $M_t = M(X_t)$, $b_t = b(X_t)$, $f_{e,t} = f_e(0|X_t)$ is the conditional density of $e_t = b_t(\varepsilon_t - \eta_0)$ evaluated at 0 conditional on X_t, $B_{LC,t} = \mu_2[(1/2)q_t'' f_t + q_t' f_t']/f_t$, $\mu_2 = \int K(v)v^2 dv$, $f_t = f(X_t)$, $q_t'' = [d^2 q_\tau(x)/dx^2]|_{x=X_t}$, q_t' and f_t' are the first derivatives of $q_\tau(\cdot)$ and $f(\cdot)$ evaluated at X_t, respectively.

Proposition 2. *Under the same conditions as in Proposition 2.1, we have*

$$\hat{h} = (C_2/4C_3)^{1/5} n^{-1/5} + o_p(n^{-1/5}),$$

where

$$C_2 = (1/2)\nu_0 E[M_t f_{e,t} b_t^2 / f_t^2],$$

$$C_3 = (1/2)E\left[M_t f_{e,t} B_{LC,t}^2\right] - (1/2)\left(E[M_t f_{e,t} B_{LC,t} b_t]\right)^2 E[M_t f_{e,t} b_t^2],$$

and

$$\nu_0 = \int K^2(v) dv.$$

The proofs of Proposition 1 and Proposition 2 are in Appendix 2.

Proposition 1 implies hat $\hat{\delta} - \delta_0 = O_p(h^2 + n^{-1/2})$, by Taylor expansion argument. Proposition 2 gives that the optimal smooth parameter in the second step has the order of $n^{-1/5}$, the same order when the data is identically independently distributed. With Proposition 1 and Proposition 2, we can derive the asymptotic distribution of τth direct nonparametric conditional quantile estimator $\hat{q}_\tau(x)$ for absolutely regular processes.

Theorem 1. *Under assumptions A.1-A.5, the asymptotic normal distribution $\hat{q}_\tau(x)$ is given* *as follows*

$$\sqrt{nh}[\hat{q}_\tau(x) - q_\tau(x) - B_1 h^2] \xrightarrow{d} N(0, V(x))$$

where

$$B_1 = B_{LC}(x) - b(x)E[M_t f_{e,t} B_{LC,t} b_t]/E[M_t f_{e,t} b_t^2],$$

$$B_{LC}(x) = \mu_2[(1/2)q''_\tau(x)f_X(x) + q'_\tau(x)f'_X(x)]/f_X(x),$$

and

$$V(x) = \nu_0 b^2(x)/f(x).$$

The proof of Theorem 1 is in Appendix 2.

Theorem 1 states that the asymptotic normal distribution of τth direct nonparametric conditional quantile estimator $\hat{q}_\tau(x)$ for the absolutely regular β-mixing processes is the same as when the data is identically independently distributed. Therefore, the efficiency gain of direct nonparametric conditional quantile estimator at extreme tail quantiles with i.i.d data remains hold when the data process is the absolutely regular β-mixing processes, with its mixing coefficient β_r satisfying assumption A.1.

4.1.2 The Local Linear Estimator

One can also estimate the conditional quantile function using the local linear estimation method. The local linear estimator has some advantages over the local constant estimator such as that it has smaller estimation bias than the local constant estimator at the boundary region of data support and that it can simultaneously estimate the conditional quantile function $q_\tau(x)$ and its derivative function $q'_\tau(x) = \frac{\partial q_\tau(x)}{dx}$. The local linear estimator $(q_\tau(x), q_\tau^{'}(x))^\top$ is given by $(\hat{q}_\tau(x), \hat{q}_\tau^{'}(x))^\top \equiv (\hat{\alpha}(x), \hat{\beta}(x))^\top$ as follows:

$$\begin{aligned}
\hat{\gamma}(x) &\overset{def}{=} \begin{pmatrix} \hat{\alpha}(x) \\ \hat{\beta}(x) \end{pmatrix} \\
&= \arg\min_{\alpha,\beta} \frac{1}{n} \sum_{t=1}^{n} \left[Q_{N(Y_t, \hat{b}(X_t)),\delta} - \alpha - (X_t - x)\beta \right]^2 K_{tx} \\
&= \left[\frac{1}{nh} \sum_{s \neq t}^{n} \tilde{X}_{st} \tilde{X}_{st}^\top K_{st} \right]^{-1} \frac{1}{nh} \sum_{s \neq t}^{n} \tilde{X}_{st} K_{st} Q_{N(Y_t, \hat{b}(X_t)),\delta} \qquad (4.3)
\end{aligned}$$

where $\tilde{X}_{st} = (1, X_s - X_t)^\top$, $K_{tx} = K((X_t - x)/h)$, the definitions of $Q_{N(Y_t, \hat{b}(X_t)),\delta}$, $\hat{b}(X_t)$ and δ are the same as defined in Section 4.1.1.

We choose (η, h) by minimizing the following cross-validation objective function

$$\frac{1}{n} \sum_{t=1}^{n} \rho_\tau(Y_t - \hat{q}_{\tau,LL,-t}(X_t)),$$

where $\rho_\tau(v) = v[1 - 1(v \leq 0)]$ is the check function, $\hat{q}_{\tau,LL,-t}(X_t) = \hat{\alpha}_{-t}(X_t) = \iota_1^\top \hat{\gamma}_{-t}(X_t)$ is the leave-one-out estimator of $q_\tau(X_t)$ with $\iota_1 = (1, 0)^\top$, and $\hat{\gamma}_{-t}(X_t)$ is given by

$$\hat{\gamma}_{-t}(X_t) = \begin{pmatrix} \hat{\alpha}_{-t}(X_t) \\ \hat{\beta}_{-t}(X_t) \end{pmatrix} = \left[\frac{1}{nh} \sum_{s \neq t}^{n} \tilde{X}_{st} \tilde{X}_{st}^\top K_{st} \right]^{-1} \frac{1}{nh} \sum_{s \neq t}^{n} \tilde{X}_{st} K_{st} Q_{N(Y_t, \hat{b}(X_t)),\delta},$$

where $\tilde{X}_{st} = (1, X_s - X_t)^\top$ and $K_{st} = K((X_s - X_t)/h)$. The asymptotic behaviour of $\hat{\eta}_{LL}$, $\hat{h}_{LL}$ and the asymptotic distribution of the direct nonparametric local linear estimator of the τth conditional quantile estimator $\hat{\gamma}(x)$ under β-mixing processes are given in the next two propositions.

Proposition 3. *Under assumption A1-A5, we have*

$$\hat{\eta}_{LL} - \eta_0 = B_{11,LL} h^2 + o_p(1)$$

$$\hat{h}_{LL} = [C_2/4C_{3,LL}]^{1/5} n^{-1/5} + o_p(1)$$

where $B_{11,LL}$ *is defined the same way as in* B_{11} *(see Proposition 1) except that* $B_{LC,t}$ *in* B_{11} *is replaced by* $B_{LL,t} = (1/2)\mu_2 q_\tau^{''}(X_t)$*,* C_2 *was defined in the Proposition 2 and* $C_{3,LL}$ *is defined the same way as* C_3 *(see Proposition 2) except that* $B_{LC,t}$ *in* C_3 *is replaced by* $B_{LL,t}$*.*

Proposition 4. *Under assumptions A1-A5, we have*

$$\sqrt{nh} D_h(\hat{\gamma}(x) - \gamma(x) - [B_{LL}(x) - B_{11,LL} m_{LL}(x)]h^2) \xrightarrow{d} N(0_2, \Sigma(x))$$

We define D_h*,* $B_{LL}(x)$*,* $m_{LL}(x)$ *and* $\Sigma(x)$ *as following:*

$D_h = (\begin{matrix} 1 & 0 \\ 0 & h \end{matrix})$

$B_{LL}(x) = (B_{LL,1}(x), B_{LL,2}(x))^\top$ *where* $B_{LL,1}(x) = (1/2)\mu_2 q_\tau''(x)$ *and* $B_{LL,2}(x) = (\mu_4 - \mu_2^2) f'(x) q_\tau''(x)/[2\mu_2 f(x)] + \mu_4 q_\tau'''(x)/(6\mu_2)$

$m_{LL}(x) = (b(x), \xi'(x)/f(x))^\top$ *and* $\xi(x) = b(x)f(x)$ *so that* $\xi'(x) = b'(x)f(x) + b(x)f'(x)$

$\Sigma(x) = \frac{b^2(x)}{f(x)}(\begin{matrix} v_0 & 0 \\ 0 & \frac{v_2}{\mu_2^2} \end{matrix})$ *where* $v_0 = \int K(v)dv$ *and* $v_2 = \int K^2(v)v^2 dv$

The Proposition 2.3 and 2.4 can be proved as the same way as the proof of Proposition 2.1, 2.2 and Theorem 2.1, but by utilizing the nonparametric local linear estimation arguments as in the proof of Proposition 4.1-4.3 in [65]. Note that as in the i.i.d case, the nonparametric direct local linear estimator of conditional quantile under β-mixing processes and the nonparametric direct local constant estimator have the same asymptotic variance. Therefore, the property of efficiency gain at extreme tail quantiles of direct nonparametric local constant estimator of conditional quantile compared to check function approach sustains in the local linear estimation approach.

4.2 Simulation

In this section we present a Monte Carlo simulation study for examining the finite sample performance of our proposed direct conditional quantile estimator. We show both local constant and local linear estimation results of the proposed direct method results ('direct lc' and 'direct ll') and compare them with the traditional check function method counterparts ('check lc' and 'check ll'). For the check function based method, we select the bandwidths by minimizing the value of delete-one cross-validated check function.

We consider three location-scale data generating processes (DGPs). The dependent variable Y_t follows a nonlinear AR(1) process. It has location $a(Y_{t-1}) = 0.9sin(2\pi Y_{t-1})$, constant scale $b(Y_{t-1}) = 1$, and error terms are Gaussian, standardized Student-t_v, and χ_v^2. We de-mean all the distributions. We set the variance of Gaussian to be two and the degrees of freedom of each Student-t_v and χ_v^2 to be five. Thus, the data is generated as follows:

$$Y_t = 0.9 * sin(2\pi Y_{t-1}) + u_t \tag{4.4}$$

where $Y_0 = 0$ and u_t is i.i.d $N(0, 2)$, or $t_{v=5}$, or $\chi^2_{v=5}$.

We use the Gaussian kernel function for the estimation. We replicate $M = 1000$ Monte Carlo simulations for each of the sample sizes $n = (100, 200, 400)$. We report the median values of MSE for $\tau = (0.01, 0.05, 0.25, 0.5, 0.75, 0.95, 0.99)$ in tables 4.1-4.3. The MSE is computed by equation (4.5) where $\hat{q}_\tau(Y_t|Y_{t-1})$ is the pointwise estimated conditional quantile of Y_t given Y_{t-1} and $q_\tau(Y_t|Y_{t-1})$ is the counterpart true value.

$$MSE = \frac{1}{n}\sum_{t=2}^{n}(\hat{q}_\tau(Y_t|Y_{t-1}) - q_\tau(Y_t|Y_{t-1}))^2. \tag{4.5}$$

The simulation results for Gaussian distributed error terms are reported in Table 4.1. The results show that the proposed direct method is not only more efficient than the traditional check method at the extreme quantiles as expected from our theoretical propositions but also outperform it at the middle quantiles. The results are similar to [65] where they show for iid data, their proposed method has smaller finite sample estimation MSE for normally distributed ϵ.

Table 4.2 and Table 4.3 give estimation results for Student-t_5 and for χ^2_5 distributed ϵ. We observe that our proposed estimator have smaller finite sample median MSE for most cases than that of the check function based estimators.

Table 4.1: Median 10^{-2} MSE Values Normal Distribution

Estimator	n	$\tau = 0.01$	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$	$\tau = 0.99$
direct lc	100	41.24	21.39	14.49	12.4	14.03	23.57	40.23
check lc	100	56.27	28.12	17.3	14.61	16.16	28.58	56.06
direct ll	100	37.98	19.71	12.66	11.72	12.19	19.66	36.73
check ll	100	68.78	32.82	16.82	14.6	16.66	33.95	69.25
direct lc	200	23.64	14.8	8.61	7.59	8.51	14.35	24.97
check lc	200	37.51	18.51	10.48	9.27	10.2	18.03	37.71
direct ll	200	20.4	11.81	7.64	6.74	7.55	11.53	21.4
check ll	200	44.37	20.43	9.92	8.66	10.01	20.36	45.65
direct lc	400	15.51	9.47	5.06	4.36	5.17	9.09	14.96
check lc	400	23.15	12.39	6.39	5.77	6.51	12.09	24.74
direct ll	400	11.88	7.13	4.85	4.47	4.81	7.19	11.88
check ll	400	27.54	12.13	6.24	5.43	6.23	12.53	28.83

In the next section we will compare forecasting performances of our estimator and the check function based estimator using U.S housing returns data.

Table 4.2: Median MSE Values in 10^{-2} for Student-t Dist.

Estimator	n	$\tau = 0.01$	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$	$\tau = 0.99$
direct lc	100	68.94	24.6	11.84	9.38	11.68	24.36	68.1
check lc	100	109.95	32.86	13.06	10.73	13.45	33.13	107.54
direct ll	100	67.97	21.06	9.22	8.31	9.56	19.81	62.17
check ll	100	125.01	38.13	11.72	9.12	11.66	38.6	125.15
direct lc	200	41.92	16.93	7.7	5.82	7.75	16.27	41.5
check lc	200	75.03	21.72	8.86	6.82	8.7	21.39	72.64
direct ll	200	39.67	12.46	5.8	5.2	5.88	12.93	39.04
check ll	200	94.44	24.27	7.41	5.8	7.25	24.06	93.64
direct lc	400	28.65	11.06	4.75	3.79	4.8	11.07	26.96
check lc	400	50.11	14.75	5.55	4.36	5.45	14.74	47.79
direct ll	400	22.84	7.91	4.08	3.51	3.99	7.5	21.67
check ll	400	64.9	15.69	4.91	3.72	4.65	15.65	61.64

Table 4.3: Median MSE Values in 10^{-2} for Standardized χ^2_5 Dist.

Estimator	n	$\tau = 0.01$	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$	$\tau = 0.99$
direct lc	100	3.9	2.22	1.79	2.27	3.9	11.82	24.19
check lc	100	4.22	2.35	2.12	3.02	4.96	14	35.48
direct ll	100	1.75	1.33	1.27	1.57	2.33	6.95	18.94
check ll	100	2.85	1.51	1.38	2.03	3.97	14.37	39.87
direct lc	200	2.54	1.31	1.06	1.25	2.27	6.58	16.25
check lc	200	2.56	1.39	1.26	1.76	3.06	8.19	23.82
direct ll	200	1.04	0.84	0.82	0.97	1.44	3.58	11.2
check ll	200	1.48	0.84	0.88	1.24	2.27	8.62	28.51
direct lc	400	1.31	0.8	0.64	0.75	1.31	3.87	9.56
check lc	400	1.44	0.83	0.76	1.06	1.7	5.33	14.71
direct ll	400	0.69	0.54	0.49	0.6	0.9	2.27	5.34
check ll	400	0.76	0.47	0.51	0.75	1.32	5.28	17.51

4.3 An Empirical Application

The 2008 economic recession that emerged from a series of subprime mortgage crisis has reflected the significant impact of housing market on financial markets and even the entire economy. Given the fact that the global economy is still struggling in recovering from last recession and owning a home being an sizable part of private households net wealth [76], better methods of forecasting the movements of housing price returns are essential in understanding the implications of housing market on economy. Even though the vast majority of the existing works focus on forecasting the mean value of housing prices with traditional parametric methods [77, 78, 79, 80], forecasting the conditional quantiles is more helpful in understanding the risks of housing market by offering a comprehensive future housing prices distribution. Since it is extremely challenging to efficiently estimate the price serial correlation of infrequent traded assets with parametric method, our proposed direct nonparametric method fits this study well.

We consider forecasting the U.S housing return quantiles. The data is from a publicly available dataset, S&P/Case-Shiller House Price Indices. It is a national representative and widely quoted data. In addition, the financial derivatives on this index have been traded at the Chicago Mercantile Exchange since May 2006. [81] thoroughly discussed the value and importance of such products.

Home Price Index covers periods from January, 1975 to December, 2015. For this study, we focus on monthly frequency data. As we can see from Figure 4.1, the housing index data fully captures the major phenomenons of U.S housing market. The housing prices were going high from the beginning of early 2000s, declined sharply during 2008 recession and recovered after 2010. We denote the housing index observations as I_t and housing return as y_t which is defined as

$$Y_t = \frac{I_t - I_{t-1}}{I_{t-1}}. \tag{4.1}$$

The dependent variable to be forecasted is the home price return, y_t, and the independent variable is one month lag of home price return, y_{t-1}. We keep a fixed window of in-sample size of 240 observations and roll the in-sample estimation window forward till the last available observation on December, 2015. The in-sample estimation starts from January, 1975 and first forecast begins at January, 1995. We forecast the quantiles at $\tau = (0.05, 0.25, 0.5, 0.75, 0.95)$ with proposed direct method by equa-

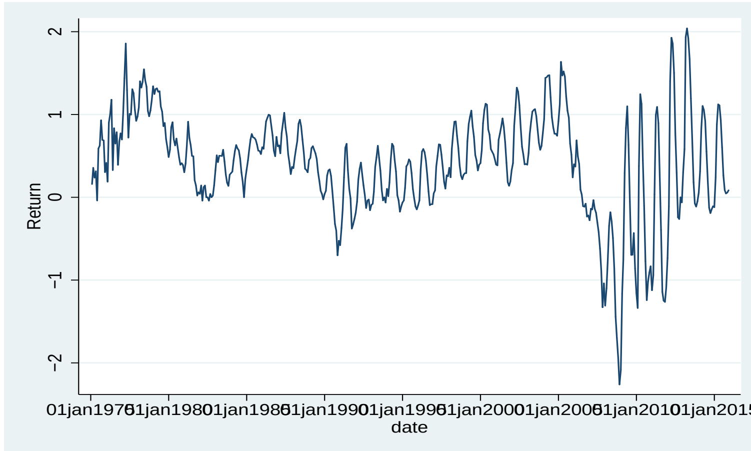

Figure 4.1: Housing Index Return

tion (4.2) for local constant version and equation (4.3) for local linear version. The δ and bandwidth are selected by minimizing the cross-validation function of equation $CV_i(\delta, h) = \frac{1}{n}\sum_{t=1+i}^{240+i} \rho_\tau(Y_t - \hat{q}_{\delta,h,-1}(Y_{t-1}))M(Y_{t-1})$ with $n = 240$ where ρ_τ is the check function $\rho_\tau(v) = v(\tau - I(v \leq 0))$ and $\hat{b}(Y_i)$ is estimated by calculating the square root of $E(Y_t^2|Y_{t-1}) - [E(Y_t|Y_{t-1}]^2$.

$$\hat{q}_\tau(Y_t|Y_{t-1}) = \arg\min_{\alpha} \sum_{i=t-242}^{t-2} \{Q_{N(Y_{i+1},\hat{b}(Y_i)),\delta} - \alpha\}^2 K_h(Y_i, Y_{t-1}). \tag{4.2}$$

$$\hat{q}_\tau(Y_t|Y_{t-1}) = \arg_{\alpha}\{\min_{\gamma=(\alpha,\beta)} \sum_{i=t-242}^{t-2} \{Q_{N(Y_{i+1},\hat{b}(Y_i)),\delta} - \alpha - (Y_i - Y_{t-1})\beta\}^2 K_h(Y_i, Y_{t-1})\}. \tag{4.3}$$

For illustration purpose, we compare our proposed direct quantile model with the traditional check function based model and linear conditional quantile model. We present the forecasts of conditional quantiles along with its associated check function values. The local constant check function method is forecasted by equation (4.4) and the local linear version is forecasted by equation (4.5).

$$\hat{q}_{\tau,lc}(Y_t|Y_{t-1}) = \arg\min_{\alpha} \sum_{i=t-242}^{t-2} \{\rho_\tau(Y_{i+1} - \alpha)K_h(Y_i, Y_{t-1})\}, \tag{4.4}$$

$$\hat{q}_{\tau,ll}(Y_t|Y_{t-1}) = \arg_{\alpha}\{\min_{\gamma=(\alpha,\beta)} \sum_{i=t-242}^{t-2} \{\rho_\tau(Y_{i+1} - \alpha - (Y_i - Y_{t-1})\beta))K_h(Y_i, Y_{t-1})\} \tag{4.5}$$

First, we look at the conditional quantiles forecasts of local constant version at figure 4.2 for check function method approach and at figure 4.3 for direct method approach. The diamond line in the figures are housing returns realizations and the solid lines in different colors are for quantiles of various τ. We can observe from the figures that the direct method results cover the housing return realizations much better than the check function results.

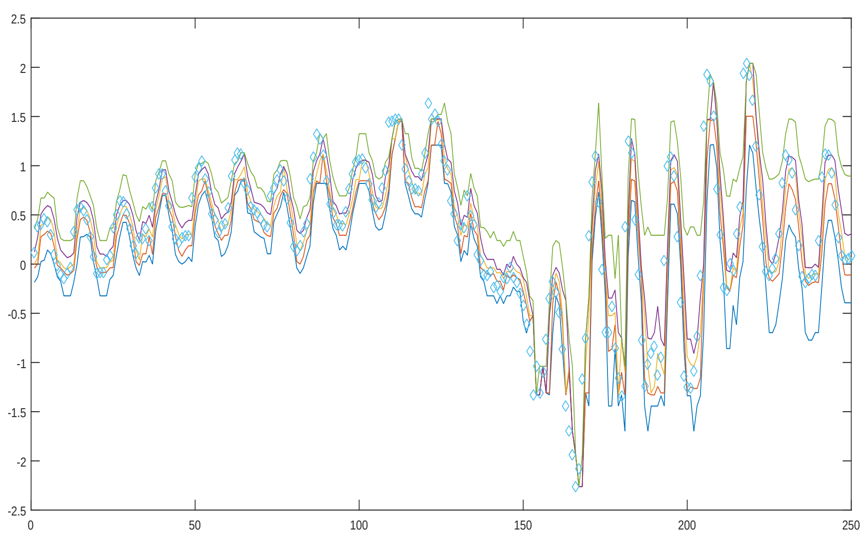

Figure 4.2: Check Function Method-Local Constant

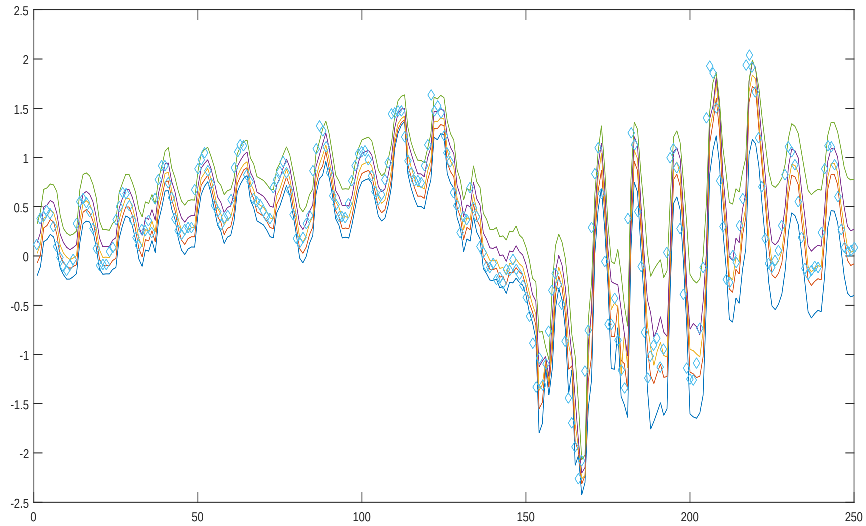

Figure 4.3: Direct Method-Local Constant

Again, we look at the forecasts of local-linear version at figure 4.4 for check function method approach and at figure 4.5 for direct method approach. The results are persistent in the pattern that direct method covers the housing return realizations better than the check function based method.

The performance of out-of-sample forecasting is evaluated based on the check function values. The smaller check function values are, the better forecasting results are. We report the check function values of direct method, check function based method, and linear method on Table 4.4. As table shows, our proposed direct methods delivers a much more accurate forecasting results in terms of check function values than conventional check function method and linear method. In tail regions ($\tau = 0.05$ and $\tau = 0.95$), the direct results are improved roughly 10% and 5.7% for local constant estimator and 5.3% and 23.3% for local linear estimator comparing to check function method and almost 50% for both versions of linear method.

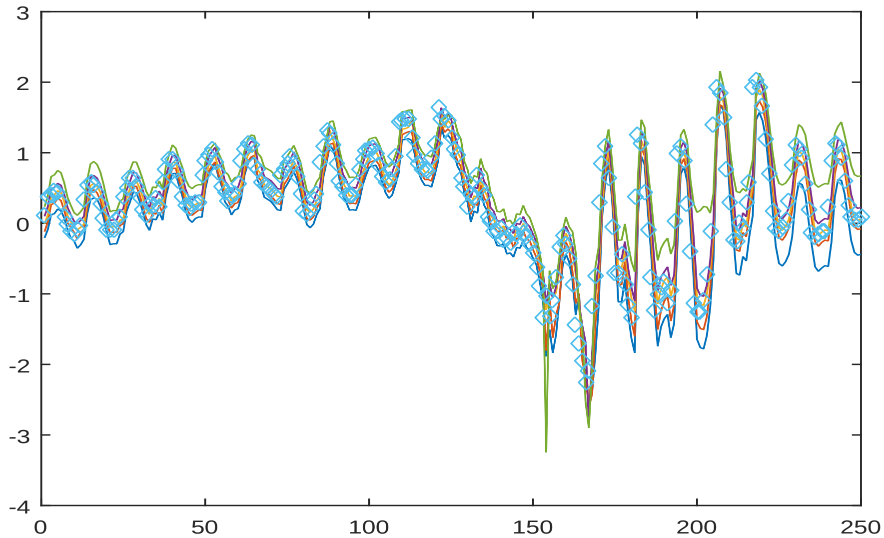

Figure 4.4: Check Function Method-Local Linear

Table 4.4: Window Size: 240; Number of Forecasting: 250

Estimator	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
direct lc	8.96	22.65	28.39	26.35	12.65
check lc	9.95	22.54	28.57	25.84	13.41
direct ll	8.88	22.32	27.28	25.21	12.20
check ll	9.39	22.54	27.85	25.05	15.92
linear	22.89	25.12	27.20	27.99	27.35

For robustness check purpose, we repeat the entire process with an smaller window; a fixed window of in-sample size of 120 observations. We report the results on Table 4.5. We found that the result patterns described above are persistent for different forecasting window sizes.

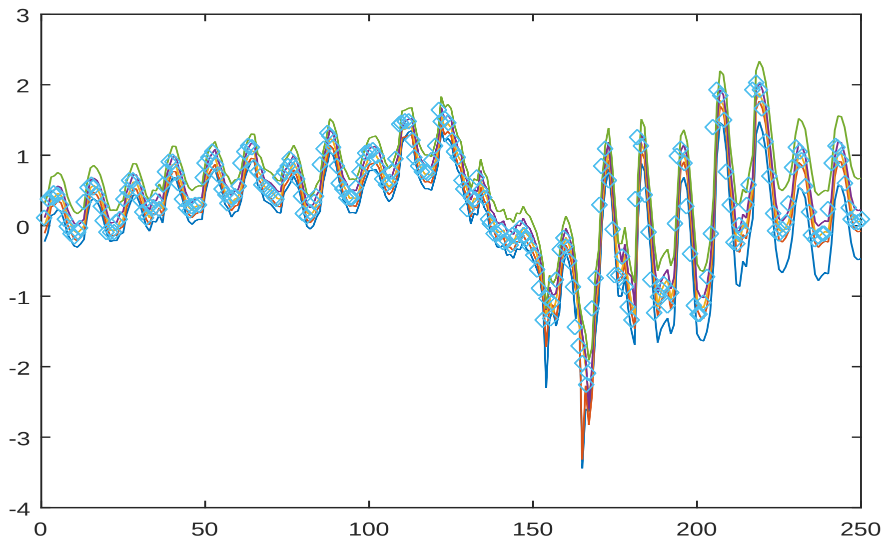

Figure 4.5: Direct Method-Local Linear

Table 4.5: Window Size: 120; Number of Forecasting: 370

Estimator	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
direct lc	10.30	28.54	36.36	32.92	14.24
check lc	12.00	28.89	37.24	32.49	14.99
direct ll	9.71	27.44	35.19	31.54	15.94
check ll	10.73	27.71	36.24	31.41	16.16
linear	26.68	31.22	34.46	35.32	33.94

5. CONCLUSION

In the first essay, we study the housing vacancy rate. It's challenging to estimate the housing vacancy rate for the transitional economies since not much historical survey data is available at developing countries. We propose a novel identification strategy to estimate the housing vacancy rate by transaction data, which is housing units level electricity data in this paper. We estimate the housing vacancy rate to be 15.52% showing the high volumes of vacant properties among Chinese housing market. These vacant properties are more likely for investment purpose than for housing services since the transaction cost of vacant properties are much lower than that of occupied ones. Therefore, the vacant houses are more liquid on the market. Moreover, the large price drops will put further pressure on the downward pricing trend and eventually affect other sectors of the economy.

In the second essay, we study the phenomenon of demand stickiness. We set up a structural model to describe the retailer's profit maximizing process when the consumers face a re-optimization cost. We propose a method to estimate the magnitude of the demand stickiness. Although consumer rational inattention has been theorized in the literature [29, 30], as far as we know, there is no study that has empirically estimated the level of consumer rational inattention. Using Dominick's scanner data, our estimation confirms the existence of demand stickiness, and moreover, we find that for typical consumers, their re-optimization thresholds are about 10% of the average product price. In addition, we also examine the effect of demographics on the threshold on demand stickiness. Our results show that the re-optimizing threshold is positively correlated with the proportion of hurried shoppers and negatively correlated with the fraction of shoppers with college degree or above and the average household size. These results support our logic based on consumer re-optimization cost. In addition, the re-optimizing threshold does not seem to vary among stores of different price tiers, or between private labels and national brands.

An important application of demand stickiness lies in dynamic pricing. We propose a new dynamic pricing strategy that accommodates sticky demand. We conduct a counterfactual analysis and show that a retailer could see a 1% increase in revenue by taking into account demand stickiness when dynamically setting prices to clear its inventory. Given the low margin in the grocery retail industry, this increase in revenue could be significant for the retailer. In addition, the revenues generated by our new dynamic pricing strategy are more stable than those generated by the traditional dynamic pricing, which is important for risk-averse retailers.

In the third essay, we extend the novel method of conditional quantile estimator of [65] to weakly dependent data case so that applied researchers can apply it to estimate and forecast conditional quantile functions with time series data. We prove the consistency and asymptotic normality of our direct nonparametric conditional quantile estimator with β-mixing data. We show that the efficiency of the direct method by Monte Carlo simulations and illustrate how to apply the method on empirical data to forecast U.S. housing index. The proposed method can be widely used in forecasting economics/financial variables and in risk management analysis. With more efficient and accurate forecasting of financial variable quantile distributions, researchers may have a comprehensive understanding of financial variables for future periods.

The proposed estimation method can be further generalized to estimate semiparametric conditional quantile functions such as estimating a partially linear or a varying coefficient conditional quantile functions. One can also develop model specification tests for parametric conditional quantile functions. Since the proposed method estimate the tail distributions more accurately than the conventional check function based method, we expect test statistics based on the new method have better finite sample performances especially when testing tail part of a parametric conditional distribution. We leave the investigation of these problems as possible future research topics.

REFERENCES

[1] K. T. Rosen and L. B. Smith, "The price-adjustment process for rental housing and the natural vacancy rate," *The American Economic Review*, vol. 73, no. 4, pp. 779–786, 1983.

[2] W. C. Wheaton, "Vacancy, search, and prices in a housing market matching model," *Journal of Political Economy*, pp. 1270–1292, 1990.

[3] C. Rapaport, "Rent regulation and housing-market dynamics," *The American Economic Review*, vol. 82, no. 2, pp. 446–451, 1992.

[4] E. Belsky, J. L. Goodman, *et al.*, "Explaining the vacancy rate-rent paradox of the 1980s.," *Journal of Real Estate Research*, vol. 11, no. 3, 1996.

[5] S. A. Gabriel and F. E. Nothaft, "Rental housing markets, the incidence and duration of vacancy, and the natural vacancy rate," *Journal of Urban Economics*, vol. 49, no. 1, pp. 121–149, 2001.

[6] J. Halket and M. P. M. di Custoza, "Homeownership and the scarcity of rentals," *Journal of Monetary Economics*, vol. 76, pp. 107–123, 2015.

[7] P. Cheshire, C. A. Hilber, and H. R. Koster, "Regulating housing vacancies away? the paradoxical effects of mismatch," *Spatial Economics Research Centre*, 2015.

[8] J. Zabel, "A dynamic model of the housing market: The role of vacancies," *The Journal of Real Estate Finance and Economics*, vol. 53, no. 3, pp. 368–391, 2016.

[9] J. Wu, Y. Deng, and H. Liu, "House price index construction in the nascent housing market: the case of China," *The Journal of Real Estate Finance and Economics*, vol. 48, no. 3, pp. 522–545, 2014.

[10] L. Gan, "Research report of China household finance survey 2012," tech. rep., Southwestern University of Finance and Economics Press, 2012.

[11] H. Fang, Q. Gu, W. Xiong, and L.-A. Zhou, "Demystifying the chinese housing boom," in *NBER Macroeconomics Annual 2015, Volume 30*, University of Chicago Press, 2015.

[12] K. Chen and Y. Wen, "The great housing boom of China," *FRB St Louis Paper No. FEDLWP2014-022*, 2014.

[13] J. Wu, J. Gyourko, and Y. Deng, "Evaluating the risk of chinese housing markets: what we know and what we need to know," tech. rep., National Bureau of Economic Research, 2015.

[14] J. V. Henderson, Y. M. Ioannides, *et al.*, "A model of housing tenure choice," *American Economic Review*, vol. 73, no. 1, pp. 98–113, 1983.

[15] M. Piazzesi and M. Schneider, "Momentum traders in the housing market: survey evidence and a search model," tech. rep., National Bureau of Economic Research, 2009.

[16] W. E. Oates, "The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the tiebout hypothesis," *Journal of Political Economy*, vol. 77, no. 6, pp. 957–971, 1969.

[17] R. Dusansky, M. Ingber, and N. Karatjas, "The impact of property taxation on housing values and rents," *Journal of Urban Economics*, vol. 10, no. 2, pp. 240–255, 1981.

[18] O. Palmon and B. A. Smith, "New evidence on property tax capitalization," *Journal of Political Economy*, vol. 106, no. 5, pp. 1099–1111, 1998.

[19] K. L. Bradbury, C. J. Mayer, and K. E. Case, "Property tax limits, local fiscal behavior, and property values: Evidence from massachusetts under proposition 212," *Journal of Public Economics*, vol. 80, no. 2, pp. 287–311, 2001.

[20] S. Ménard, "Should we tax vacant dwellings? a search equilibrium model applied to the rental housing market," *Economics Letters*, vol. 117, no. 1, pp. 88–90, 2012.

[21] J. Cao and W. Hu, "A microsimulation of property tax policy in China," *Journal of Housing Economics*, vol. 33, pp. 128–142, 2016.

[22] C. Bai, Q. Li, and M. Ouyang, "Property taxes and home prices: A tale of two cities," *Journal of Econometrics*, vol. 180, no. 1, pp. 1–15, 2014.

[23] Z. Du and L. Zhang, "Home-purchase restriction, property tax and housing price in China: A counterfactual analysis," *Journal of Econometrics*, vol. 188, no. 2, pp. 558–568, 2015.

[24] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307–333, 1989.

[25] Q. Liu and G. J. Van Ryzin, "Strategic capacity rationing to induce early purchases," *Management Science*, vol. 54, no. 6, pp. 1115–1131, 2008.

[26] D. Zhang and W. L. Cooper, "Managing clearance sales in the presence of strategic customers," *Production and Operations Management*, vol. 17, no. 4, pp. 416–431, 2008.

[27] Y. Aviv and A. Pazgal, "Optimal pricing of seasonal products in the presence of forward-looking consumers," *Manufacturing & Service Operations Management*, vol. 10, no. 3, pp. 339–359, 2008.

[28] G. Gallego and R. Phillips, "Revenue management of flexible products," *Manufacturing & Service Operations Management*, vol. 6, no. 4, pp. 321–337, 2004.

[29] R. Reis, “Inattentive consumers,” *Journal of monetary Economics*, vol. 53, no. 8, pp. 1761–1800, 2006.

[30] C. A. Sims, “Implications of rational inattention,” *Journal of monetary Economics*, vol. 50, no. 3, pp. 665–690, 2003.

[31] H. A. Chen, D. Levy, S. Ray, and M. Bergen, “Asymmetric price adjustment in the small,” *Journal of Monetary Economics*, vol. 55, no. 4, pp. 728–737, 2008.

[32] M. Chen and Z.-L. Chen, “Recent developments in dynamic pricing research: multiple products, competition, and limited demand information,” *Production and Operations Management*, vol. 24, no. 5, pp. 704–731, 2015.

[33] Z.-J. M. Shen and X. Su, “Customer behavior modeling in revenue management and auctions: A review and new research opportunities,” *Production and operations management*, vol. 16, no. 6, pp. 713–728, 2007.

[34] I. Popescu and Y. Wu, “Dynamic pricing strategies with reference effects,” *Operations Research*, vol. 55, no. 3, pp. 413–429, 2007.

[35] X. Su, “A model of consumer inertia with applications to dynamic pricing,” *Production and Operations Management*, vol. 18, no. 4, pp. 365–380, 2009.

[36] J. Nasiry and I. Popescu, “Advance selling when consumers regret,” *Management Science*, vol. 58, no. 6, pp. 1160–1177, 2012.

[37] Ö. Özer and Y. Zheng, “Markdown or everyday low price? the role of behavioral motives,” *Management Science*, vol. 62, no. 2, pp. 326–346, 2015.

[38] R. Swinney, “Selling to strategic consumers when product value is uncertain: The value of matching supply and demand,” *Management Science*, vol. 57, no. 10, pp. 1737–1751, 2011.

[39] G. P. Cachon and R. Swinney, “Purchasing, pricing, and quick response in the presence of strategic consumers,” *Management Science*, vol. 55, no. 3, pp. 497–511, 2009.

[40] W. Elmaghraby and P. Keskinocak, “Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions,” *Management science*, vol. 49, no. 10, pp. 1287–1309, 2003.

[41] M. E. Slade *et al.*, “Optimal pricing with costly adjustment: evidence from retail-grocery prices,” *The Review of Economic Studies*, vol. 65, no. 1, pp. 87–107, 1998.

[42] D. Levy, H. A. Chen, G. Müller, S. Dutta, and M. Bergen, “Holiday price rigidity and cost of price adjustment,” *Economica*, vol. 77, no. 305, pp. 172–198, 2010.

[43] J. A. Chevalier, A. K. Kashyap, and P. E. Rossi, “Why don’t prices rise during periods of peak demand? evidence from scanner data,” *The American Economic Review*, vol. 93, no. 1, pp. 15–37, 2003.

[44] E. J. Warner and R. B. Barsky, “The timing and magnitude of retail store markdowns: evidence from weekends and holidays,” *The Quarterly Journal of Economics*, pp. 321–352, 1995.

[45] V. Shankar and R. N. Bolton, “An empirical analysis of determinants of retailer pricing strategy,” *Marketing Science*, vol. 23, no. 1, pp. 28–49, 2004.

[46] V. R. Nijs, S. Srinivasan, and K. Pauwels, “Retail-price drivers and retailer profits,” *Marketing Science*, vol. 26, no. 4, pp. 473–487, 2007.

[47] H. Chen, H. Marmorstein, M. Tsiros, and A. R. Rao, “When more is less: The impact of base value neglect on consumer preferences for bonus packs over price discounts,” *Journal of Marketing*, vol. 76, no. 4, pp. 64–77, 2012.

[48] S. Ray, C. A. Wood, and P. R. Messinger, "Multicomponent systems pricing: Rational inattention and downward rigidities," *Journal of Marketing*, vol. 76, no. 5, pp. 1–17, 2012.

[49] D. Levy, M. Bergen, S. Dutta, and R. Venable, "The magnitude of menu costs: direct evidence from large us supermarket chains," *The Quarterly Journal of Economics*, pp. 791–825, 1997.

[50] N. G. Mankiw, "Small menu costs and large business cycles: A macroeconomic model of monopoly," *The Quarterly Journal of Economics*, vol. 100, no. 2, pp. 529–537, 1985.

[51] D. Levy, D. Lee, H. Chen, R. J. Kauffman, and M. Bergen, "Price points and price rigidity," *Review of Economics and Statistics*, vol. 93, no. 4, pp. 1417–1431, 2011.

[52] S. Peltzman, "Prices rise faster than they fall," *Journal of political economy*, vol. 108, no. 3, pp. 466–502, 2000.

[53] E. T. Anderson and D. I. Simester, "The role of sale signs," *Marketing Science*, vol. 17, no. 2, pp. 139–155, 1998.

[54] R. B. Barsky, M. Bergen, S. Dutta, and D. Levy, "What can the price gap between branded and private-label products tell us about markups?," in *Scanner data and price indexes*, pp. 165–228, University of Chicago Press, 2003.

[55] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

[56] M. Jones and P. Hall, "Mean squared error properties of kernel estimates or regression quantiles," *Statistics & Probability Letters*, vol. 10, no. 4, pp. 283–289, 1990.

[57] P. Chaudhuri *et al.*, "Nonparametric estimates of regression quantiles and their local bahadur representation," *The Annals of statistics*, vol. 19, no. 2, pp. 760–777, 1991.

[58] K. Yu and M. Jones, "Local linear quantile regression," *Journal of the American statistical Association*, vol. 93, no. 441, pp. 228–237, 1998.

[59] T. Honda, "Nonparametric estimation of a conditional quantile for α-mixing processes," *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 3, pp. 459–470, 2000.

[60] M. Jones and K. Yu, "Improved double kernel local linear quantile regression," *Statistical Modelling*, vol. 7, no. 4, pp. 377–389, 2007.

[61] Q. Li and J. S. Racine, "Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data," *Journal of Business & Economic Statistics*, vol. 26, no. 4, pp. 423–434, 2008.

[62] B. Kai, R. Li, and H. Zou, "Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 1, pp. 49–69, 2010.

[63] V. Spokoiny, W. Wang, and W. K. Härdle, "Local quantile regression," *Journal of Statistical Planning and Inference*, vol. 143, no. 7, pp. 1109–1129, 2013.

[64] A. Daouia, L. Gardes, S. Girard, *et al.*, "On kernel smoothing for extremal quantile regression," *Bernoulli*, vol. 19, no. 5B, pp. 2557–2589, 2013.

[65] J. S. Racine and K. Li, "Direct nonparametric conditional quantile estimation," *forthcoming in Journal of Econometrics*, 2016.

[66] S. A. Khashimov, "On the asymptotic distribution of the generalized u-statistics for dependent variables," *Theory of Probability & Its Applications*, vol. 32, no. 2, pp. 373–375, 1988.

[67] S. A. Khashimov, "Limiting behavior of generalized u-statistics of weakly dependent stationary processes," *Theory of Probability & Its Applications*, vol. 37, no. 1, pp. 148–150, 1993.

[68] K.-i. Yoshihara *et al.*, "Limiting behavior of generalized quadratic forms generated by absolutely regular sequences ii," 1990.

[69] K.-i. Yoshihara *et al.*, "Limiting bahavior of generalized quadratic forms generated by regular sequences iii," 1992.

[70] H. Takahara, K.-i. Yoshihara, *et al.*, "Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences," 1987.

[71] Y. Fan and Q. Li, "Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing," *Journal of Nonparametric Statistics*, vol. 10, no. 3, pp. 245–271, 1999.

[72] Y. K. Truong and C. J. Stone, "Semiparametric time series regression," *Journal of Time Series Analysis*, vol. 15, no. 4, pp. 405–428, 1994.

[73] D. W. Andrews, "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica: Journal of the Econometric Society*, pp. 43–72, 1994.

[74] Y. Fan and Q. Li, "Root-n-consistent estimation of partially linear time series models," *Journal of Nonparametric Statistics*, vol. 11, no. 1-3, pp. 251–269, 1999.

[75] P. M. Robinson, "Root-n-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, pp. 931–954, 1988.

[76] M. Bertus, H. Hollans, and S. Swidler, "Hedging house price risk with cme futures contracts: the case of las vegas residential real estate," *The Journal of Real Estate Finance and Economics*, vol. 37, no. 3, pp. 265–279, 2008.

[77] K. E. Case and R. J. Shiller, "The efficiency of the market for single-family homes," *American economic review*, pp. 125–137, 1989.

[78] K. E. Case and R. J. Shiller, "Forecasting prices and excess returns in the housing market," *Real Estate Economics*, vol. 18, no. 3, pp. 253–273, 1990.

[79] W. Miles, "Boom–bust cycles and the forecasting performance of linear and non-linear models of house prices," *The Journal of Real Estate Finance and Economics*, vol. 36, no. 3, pp. 249–264, 2008.

[80] A. Gu, "The predictability of house prices," *Journal of Real Estate Research*, 2009.

[81] R. J. Shiller, "Derivatives markets for home prices," tech. rep., National Bureau of Economic Research, 2008.

APPENDIX A

APPENDIX FOR CHAPTER 2

Proposition. *The profit maximizer p_t^* for Equation (3.7) is contained by the set of* $\{p_{t-1} + \frac{A-\sqrt{A^2-B}}{3\beta_1\gamma}, p_{t-1} - \frac{-A-\sqrt{A^2+B}}{3\beta_1\gamma}, \frac{c_t-\beta_0/\beta_1}{2}, p_{t-1}-1/\gamma, p_{t-1}+1/\gamma\} \equiv \{p_t^1, p_t^2, p_t^3, p_{t-1}-1/\gamma, p_{t-1}+1/\gamma\}$.

Proof. First, we want to show that if $p_t^1 \in (p_{t-1}, p_{t-1}+1/\gamma)$, then p_t^1 is the local maximizer on $(p_{t-1}, p_{t-1}+1/\gamma)$. Consider $p_t \in (p_{t-1}, p_{t-1}+1/\gamma)$. The first-order derivative of $\pi(p_t)$ is given by

$$\pi'(p_t) = \beta_0 + \beta_1(p_{t-1} + \gamma(p_t - p_{t-1})^2) + 2\beta_1\gamma(p_t - c_t)(p_t - p_{t-1}) \tag{4.1}$$

The equation $\pi'(p_t) = 0$ has two roots: $x_1 = p_{t-1} + \frac{A-\sqrt{A^2-B}}{3\beta_1\gamma}$ and $x_2 = p_{t-1} + \frac{A+\sqrt{A^2-B}}{3\beta_1\gamma}$. Note that $B < 0$, so $\sqrt{A^2-B} > |A|$. We have $x_1 > p_{t-1}$ and $x_2 < p_{t-1}$. The right derivative of $\pi(p_t)$ at p_{t-1} is given by

$$\lim_{p_t \to p_{t-1}+} \pi'(p_t) = \beta_0 + \beta_1 p_{t-1} = D(p_{t-1}) > 0 \tag{4.2}$$

Since $\pi(p_t)$ is a cubic form in p_t, we know that x_1 is the local maximizer, while x_2 is the local minimizer. Note that x_1 is p_t^1, so if $p_t^1 \in (p_{t-1}, p_{t-1}+1/\gamma)$, then p_t^1 is the local maximizer on $(p_{t-1}, p_{t-1}+1/\gamma)$.

$$\pi(p_t^1) > \pi(p_t), \quad \forall p_t \in (p_{t-1}, p_{t-1}+1/\gamma) \tag{4.3}$$

If $p_t^1 > p_{t-1} + 1/\gamma$, then $\pi'(p_t) > 0$ on $(p_{t-1}, p_{t-1} + 1/\gamma)$.

$$\pi(p_{t-1} + 1/\gamma) > \pi(p_t), \quad \forall p_t \in (p_{t-1}, p_{t-1} + 1/\gamma) \tag{4.4}$$

Second, we want to show that if $p_t^2 \in (p_{t-1} - 1/\gamma, p_{t-1})$ and $p_{t-1} > c_t$, then p_t^2 is the local maximizer on $(p_{t-1} - 1/\gamma, p_{t-1})$. Consider $p_t \in (p_{t-1} - 1/\gamma, p_{t-1})$. The first-order derivative of $\pi(p_t)$ is given by

$$\pi'(p_t) = \beta_0 + \beta_1(p_{t-1} - \gamma(p_t - p_{t-1})^2) - 2\beta_1\gamma(p_t - c_t)(p_t - p_{t-1}) \tag{4.5}$$

The equation $\pi'(p_t) = 0$ has two roots: $x_1 = p_{t-1} - \frac{-A-\sqrt{A^2+B}}{3\beta_1\gamma}$ and $x_2 = p_{t-1} - \frac{-A+\sqrt{A^2+B}}{3\beta_1\gamma}$. Note that $B < 0$, so $\sqrt{A^2 - B} < |A|$. Since $p_{t-1} > c_t$, we have $A > 0$. Thus, $x_1 < x_2 < p_{t-1}$. The left derivative of $\pi(p_t)$ at p_{t-1} is given by

$$\lim_{p_t \to p_{t-1}-} \pi'(p_t) = \beta_0 + \beta_1 p_{t-1} = D(p_{t-1}) > 0 \tag{4.6}$$

Since $\pi(p_t)$ is a cubic form in p_t, we know that x_1 is the local maximizer, while x_2 is the local minimizer. Note that x_1 is p_t^2, so if $p_t^2 \in (p_{t-1} - 1/\gamma, p_{t-1})$, then p_t^2 is the local maximizer on $(p_{t-1} - 1/\gamma, p_{t-1})$.

$$\pi(p_t^2) > \pi(p_t), \quad \forall p_t \in (p_{t-1} - 1/\gamma, p_{t-1}) \tag{4.7}$$

If $p_t^2 < p_{t-1} - 1/\gamma$, then

$$max\{\pi(p_{t-1} - 1/\gamma), \pi(p_{t-1})\} > \pi(p_t), \quad \forall p_t \in (p_{t-1} - 1/\gamma, p_{t-1}) \tag{4.8}$$

Third, we consider $p_t \in [c_t, p_{t-1} - 1/\gamma] \cup [p_{t-1} + 1/\gamma, -\beta_0/\beta_1]$. In this case, the profit function reduces to the following.

$$\pi(p_t) = (p_t - c_t)(\beta_0 + \beta_1 p_t) \tag{4.9}$$

The first order condition is given by

$$\pi'(p_t) = \beta_0 + \beta_1 p_t + \beta_1(p_t - c_t) \tag{4.10}$$

The solution to $\pi'(p_t) = 0$ is $x = \frac{c_t - \beta_0/\beta_1}{2} = p_t^3$. If $p_t^3 \in [c_t, p_{t-1} - 1/\gamma] \cup [p_{t-1} + 1/\gamma, -\beta_0/\beta_1]$, then p_t^3 is the local maximizer on $[c_t, p_{t-1} - 1/\gamma] \cup [p_{t-1} + 1/\gamma, -\beta_0/\beta_1]$. If $p_t^3 \in (p_{t-1} - 1/\gamma, p_{t-1} + 1/\gamma)$, then

$$max\{\pi(p_{t-1} - 1/\gamma), \pi(p_{t-1} + 1/\gamma)\} > \pi(p_t), \quad \forall p_t \in [c_t, p_{t-1} - 1/\gamma] \cup [p_{t-1} + 1/\gamma, -\beta_0/\beta_1] \tag{4.11}$$

In sum, the set of possible global maximizers is $\{p_t^1, p_t^2, p_t^3, p_{t-1}, p_{t-1} - 1/\gamma, p_{t-1} + 1/\gamma\}$. Since $\lim_{p_t \to p_{t-1}-} \pi'(p_t) = \lim_{p_t \to p_{t-1}+} \pi'(p_t) > 0$, p_{t-1} cannot be the global maximizer. Thus, the global maximizer p_t^* is contained by the set of $\{p_t^1, p_t^2, p_t^3, p_{t-1} - 1/\gamma, p_{t-1} + 1/\gamma\}$ $\square$

APPENDIX B

APPENDIX FOR CHAPTER 3

We will use the following short-hand notation in this Appendix. $q_{\tau,t} = q_\tau(X_t)$, $\hat{q}_{\tau,t} = \hat{q}_\tau(X_t)$, $a_t = a(X_t)$, $b_t = b(X_t)$, $\hat{b}_t = b(X_t)$, $f_t = f(X_t)$, $\hat{f}_{-t} = \hat{f}_{-t}(X_t)$, $M_t = M(X_t)$, $\eta = Q_{N(0,1),\delta}$, $\eta_0 = Q_{N(0,1),\delta_0} = Q_{\varepsilon,\tau}$, $K_{ts} = K((X_t - X_s)/h)$.

Lemma 4.1. *Under assumptions A.1-A.5, we have* $\frac{1}{n^2h}\sum_t\sum_{s\neq t}\varepsilon_s K_{ts} = O_p\left(n^{-\frac{1}{2}}\right)$

Proof of Lemma 4.1:

$$\begin{aligned} & \frac{1}{n^2h}\sum_t\sum_{s\neq t}\varepsilon_s K_{ts} \\ = & \frac{1}{n^2h}\sum_t\sum_{s<t}(\varepsilon_s K_{ts} + \varepsilon_t K_{ts}) \\ = & \frac{1}{n^2h}\sum_t\sum_{s<t}H(W_t, W_s) \end{aligned}$$

where $H(W_t, W_s) = \varepsilon_s K_{ts} + \varepsilon_t K_{ts}$ and $W_t = (x_t, \varepsilon_t)$. Let $H_1(.), s_\delta, R_n$ be defined in the same way as in Lemma B.3 with respect to the above function $H(W_t, W_s)$. By H-decomposition, we have $\frac{1}{n^2h}\sum_t\sum_{s\neq t}\varepsilon_s K_{ts} = \frac{2}{n}\sum_t H_1(W_t) + R_n$, where $H_1(W_t) = h^{-1}\varepsilon_t\int K_{ts}f(x_s)dx_s = \epsilon_t f(X_t) + R_{1t,n}$, where $R_{1t,n} = h^{-1}\varepsilon_t\int K_{ts}f(x_s)dx_s - \epsilon_t f(X_t)$. It is easy to show that $n^{-1}\sum_{t=1}^n R_{1t,n} = o_p(n^{-1/2})$. Hence,

$$\frac{2}{n}\sum_{t=1}^n H_1(W_t) = \frac{2}{n}\sum_{t=1}^n \epsilon_t f(X_t) + o_p(n^{-1/2}) = \frac{2}{n}\sum_{t=1}^n H_{1,0}(W_t) + o_p(n^{-1/2}),$$

where $H_{1,0}(W_t) = \epsilon_t f(X_t)$. Thus, $(2/n)\sum_{t=1}^n H_{1,0}(W_t)$ is the leading term of

$$(2/n)\sum_{t=1}^n H_1(W_t).$$

Below we will show the second moment of $(1/n)\sum_{t=1}^{n} H_{1,0}(W_t) = O(n^{-1})$. This then implies that conclusion of the lemma.

$$\begin{aligned} E\left[\left(\frac{1}{n}\sum_t H_{1,0}(W_t)\right)^2\right] &= \frac{1}{n^2}\left\{\sum_t E\left[H_{1,0}^2(W_t)\right] + \sum_t\sum_{s\neq t} E\left[H_{1,0}(W_t)H_{1,0}(W_s)\right]\right\} \\ &= n^{-2}\left\{C_{1n}+C_{2n}\right\}, \end{aligned}$$

$$\begin{aligned} C_{1n} &= nh^2\int E(\varepsilon_t^2|X_t)f^2(x_t)dx_t + (s.o.), \\ &= O\left(nh^2\right), \end{aligned}$$

$$\begin{aligned} C_{2n} &= \sum_t\sum_{s\neq t}\left\{\int\int H_{1,0}(w_t)H_{1,0}(w_s)dF(w_t)dF(w_s) + CM_n^{1/(1+\delta)}\beta_{|t-s|}^{\delta/(1+\delta)}\right\} \\ &= O\left(nh^2\right), \end{aligned}$$

by Lemma B.1, where

$$M_n = \max\left\{E\left[|H_{1,0}(W_t)H_{1,0}(W_s)|^{1+\delta}\right], E\left[|H_{1,0}(\tilde{W}_t)H_{1,0}(\tilde{W}_s)|^{1+\delta}\right]\right\}.$$

One can easily show that $M_n = O\left(h^{2(1+\delta)}\right)$. Also, notice that $\sum_t\sum_{s\neq t}\beta_{|t-s|}^{\delta/(1+\delta)} = O(n)$.

Therefore, $\frac{1}{n^2h}\sum_t\sum_{s\neq t}\varepsilon_s K_{ts} = O_p\left(n^{-\frac{1}{2}}\right)$.

Lemma 4.2. *Under assumptions A.1-A.5, we have*

$A_{2n} \overset{def}{=} \frac{1}{n^2h}\sum_t\sum_{s\neq t}\varepsilon_t\varepsilon_s K_{ts} = O_p((nh^{1/2})^{-1}) = o_p\left((nh)^{-1}\right).$

Proof of Lemma 4.2: As in the proof of Theorem 3.1 of Fan and Li (1999b), one can show that $E(A_{2n}^2) = O((n^2h)^{-1})$. Hence, $A_{2n} = O_p((nh^{1/2})^{-1})$.

Lemma 4.3. *Under assumptions, we have* $\frac{1}{n^2h}\sum_t\sum_{s\neq t} M_t f_{e,t} b_t(q_{\tau,s}-q_{\tau,t})K_{ts}/f(x_t) = B_{1,0}h^2 + o_p\left(nh^{-1}\right)$*, where* $B_{1,0} = E[M_t f_{e,t} B_{LC,t}]$.

Proof of Lemma 4.3:

$$
\begin{aligned}
& E\left[\frac{1}{n^2 h}\sum_t \sum_{s\neq t} M_t f_{e,t} b_t (q_{\tau,s}-q_{\tau,t}) K_{ts}/f(x_t)\right] \\
= \; & \frac{1}{n^2 h}\sum_t \sum_{s\neq t}\left(\int\int M_t f_{e,t} b_t (q_{\tau,s}-q_{\tau,t}) K_{ts} f(x_s) dx_t dx_s + C M_n^{1/(1+\delta)} \beta_{|t-s|}^{\delta/(1+\delta)}\right) \\
= \; & \frac{1}{n^2 h}\left(n^2 B_{1,0} h^3 + O\left(n h^{\frac{2+\delta}{1+\delta}}\right)\right) \\
= \; & B_{1,0} h^2 + o\left((nh)^{-1}\right),
\end{aligned}
$$

where the second equality is derived based on Lemma B.1, and $M_n = O\left(h^{2+\delta}\right)$.

$$
\begin{aligned}
& Var\left(\frac{1}{n^2 h}\sum_t \sum_{s\neq t}(q_{\tau,s}-q_{\tau,t}) K_{ts}\right) \\
\approx \; & \frac{1}{n^4 h^2}\sum_t \sum_{s\neq t} \sum_{s'\neq t\neq s} E\left[(q_{\tau,s}-q_{\tau,t}) K_{ts} (q_{\tau,s'}-q_{\tau,t}) K_{ts'}\right] \\
+ \; & \frac{1}{n^4 h^2}\sum_t \sum_{s\neq t} \sum_{t'\neq t\neq s} E\left[(q_{\tau,s}-q_{\tau,t}) K_{ts} (q_{\tau,s}-q_{\tau,t'}) K_{t's}\right],
\end{aligned}
$$

where the notation $A_n \approx B_n$ means that $A_n = B_n(1+o_p(1))$, the term with four summation indices all different from each other in the second moment is canceled from $E\{[(n^2 h)^{-1}\sum_t \sum_{s\neq t}(q_{\tau,s}-q_{\tau,t})K_{ts}]^2\}$.

By standard kernel method calculation, we have

$$
E\left[(q_{\tau,s}-q_{\tau,t}) K_{ts} (q_{\tau,s'}-q_{\tau,t}) K_{ts'}\right] = O(h^4)
$$

and

$$
Var\left(\frac{1}{n^2 h}\sum_t \sum_{s\neq t}(q_{\tau,s}-q_{\tau,t}) K_{ts}\right) = O(h/n).
$$

Therefore, $\frac{1}{n^2h}\sum_t\sum_{s\neq t} M_t f_{e,t} b_t(q_{\tau,s}-q_{\tau,t})K_{ts}/f(x_t) = B_{1,0}h^2 + o_p\left((nh)^{-1}\right)$.

Lemma 4.4. *Under assumptions A.1-A.5, we have* $\frac{1}{n^3h^2}\sum_t\sum_{s\neq t} M_t f_{e,t} b_s^2 \varepsilon_s^2 K_{ts}^2/f_t^2 = \frac{2C_2}{nh} + o_p\left((nh)^{-1}\right)$.

Proof of Lemma 4.4:

$$
\begin{aligned}
& E\left|\frac{1}{n^3h^2}\sum_t\sum_{s\neq t} M_t f_{e,t} b_s^2\varepsilon_s^2 K_{ts}^2/f_t^2\right| \\
= \ & \frac{1}{n^3h^2}\sum_t\sum_{s\neq t} E\left[M_t f_{e,t} b_s^2\varepsilon_s^2 K_{ts}^2/f_t^2\right] \\
= \ & \frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\left\{\int\int M_t f_{e,t} b_s^2\varepsilon_s^2 K_{ts}^2\frac{f(x_s)}{f(x_t)}dx_t dx_s + CM_n^{1/(1+\delta)}\beta_{|t-s|}^{\delta/(1+\delta)}\right\} \\
= \ & \frac{1}{n^3h^2}\left(2C_2n^2h + O\left(nh^{1/(1+\delta)}\right)\right) \\
= \ & \frac{2C_2}{nh} + o\left((nh)^{-1}\right),
\end{aligned}
$$

where C_2 is defined in Proposition 2, and $M_n = O\left(h\right)$ can be easily checked.

Lemma 4.5. *Under assumptions A.1-A.5,* $\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s}(q_{\tau,l}-q_{\tau,t})K_{tl}\varepsilon_s K_{ts} = o_p\left((nh)^{-1}\right)$

Proof of Lemma 4.5

$$
\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s}(q_{\tau,l}-q_{\tau,t})K_{tl}\varepsilon_s K_{ts} = \frac{1}{n^3h^2}\sum_{t<s<l} H(W_t, W_s, W_l)
$$

where $H(W_t, W_s, W_l)$ is a symmetric function. By H-decomposition, we have

$$
h^2\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s}(q_{\tau,l}-q_{\tau,t})K_{tl}\varepsilon_s K_{ts} = \frac{3}{n}\sum_t H_1(W_t) + R_n,
$$

where

$$H_1(W_t) = \varepsilon_t \int\int (q_{\tau,s} - q_{\tau,l}) K_{sl} K_{tl} dF(x_s) dF(x_l) + \varepsilon_t \int\int (q_{\tau,l} - q_{\tau,s}) K_{sl} K_{ts} dF(x_s) dF(x_l).$$

$$\begin{aligned} E\left[\left(\frac{1}{n}\sum_t H_1(W_t)\right)^2\right] &= \frac{1}{n^2}\left\{\sum_t E\left[H_1^2(W_t)\right] + \sum_t \sum_{s\neq t} E\left[H_1(W_t)H_1(W_s)\right]\right\} \\ &= n^{-2}\left\{C_{1n} + C_{2n}\right\} \end{aligned}$$

Based on the same argument as in Lemma A.1 and Lemma A.3, we can derive that $C_{1n} = C_{2n} = O\left(nh^8\right)$, $\frac{1}{n}\sum_t H_1(W_t) = O_p\left(n^{-\frac{1}{2}}h^4\right)$.

$$\begin{aligned} s_\delta &\leq C \sup_{t<s<l}\left\{E|(q_{\tau,l} - q_{\tau,t})K_{tl}\varepsilon_s K_{ts}|^{2+\delta}\right\}^{1/(2+\delta)} \\ &= O\left(h^{\frac{6+2\delta}{4+2\delta}}\right) \end{aligned}$$

Therefore, $R_n = O\left(n^{-1+\frac{\varepsilon}{2}}h^{\frac{6+2\delta}{4+2\delta}}\right)$. And $\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s}(q_{\tau,l} - q_{\tau,t})K_{tl}\varepsilon_s K_{ts} = O_p\left(n^{-\frac{1}{2}}h^2\right) + O_p\left(n^{-1+\frac{\varepsilon}{2}}h^{\frac{6+2\delta}{4+2\delta}-2}\right) = o_p\left((nh)^{-1}\right)$.

Lemma 4.6. *Under assumptions A.1-A.5,*

$$\begin{aligned} &\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s} M_t f_{e,t}(q_{\tau,s} - q_{\tau,t})K_{ts}(q_{\tau,l} - q_{\tau,t})K_{tl}/f_t^2 \\ = \;& 2C_1h^4 + o_p\left((nh^{-1})\right), \end{aligned}$$

where $C_1 = (1/2)E[M_t f_{e,t} B_{LC,t}^2]$.

Proof of Lemma 4.6

$$
\begin{aligned}
& E\left[\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s} M_t f_{e,t}(q_{\tau,s}-q_{\tau,t})K_{ts}(q_{\tau,l}-q_{\tau,t})K_{tl}/f_t^2\right] \\
= \; & \frac{1}{n^3h^2}\sum_t E\left[\frac{M_t f_{e,t}}{f_t^2}E\left[\sum_{s\neq t}\sum_{l\neq t\neq s}(q_{\tau,s}-q_{\tau,t})K_{ts}(q_{\tau,l}-q_{\tau,t})K_{tl}|x_t\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
& E\left[\sum_{s\neq t}\sum_{l\neq t\neq s}(q_{\tau,s}-q_{\tau,t})K_{ts}(q_{\tau,l}-q_{\tau,t})K_{tl}|x_t\right] \\
\leq \; & \sum_{s\neq t}\sum_{l\neq t\neq s}\left\{\int\int(q_{\tau,s}-q_{\tau,t})K_{ts}(q_{\tau,l}-q_{\tau,t})K_{tl}f(x_s)f(x_l)dx_s dx_l + CM_n^{1/(1+\delta)}\beta_{|s-l|}^{\delta/(1+\delta)}\right\} \\
= \; & B_{LC,t}^2 f_t^2 n^2 h^6 + O\left(nh^{\frac{4+2\delta}{1+\delta}}\right)
\end{aligned}
$$

where $B_{LC,t}$ was defined in Proposition 1.

$$
\begin{aligned}
& E\left[\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s} M_t f_{e,t}(q_{\tau,s}-q_{\tau,t})K_{ts}(q_{\tau,l}-q_{\tau,t})K_{tl}/f_t^2\right] \\
= \; & 2C_1h^4 + O\left(n^{-1}h^{\frac{4+2\delta}{1+\delta}-2}\right) \\
= \; & 2C_1h^4 + o\left((nh)^{-1}\right).
\end{aligned}
$$

By using similar arguments as in the proof of Lemma A.4, one can easily show that the variance of $\frac{1}{n^3h^2}\sum_t\sum_{s\neq t}\sum_{l\neq t\neq s} M_t f_{e,t}(q_{\tau,s}-q_{\tau,t})K_{ts}(q_{\tau,l}-q_{\tau,t})K_{tl}/f_t^2$ is a smaller order of $(nh)^{-1}$.

Proof of Proposition 2.1

Decomposing the objective function $CV(\eta,h) = \frac{1}{n}\sum_t^n \rho_\tau(Y_t - \hat{q}_{\tau,-t}(X_t))M_t$ into three terms as in Racine and Li (2016) as follows

$$
CV(\eta,h) \;=\; CV_1 + CV_{21}(\eta,h) + CV_{22}(\eta,h)
$$

where $CV_1 = \frac{1}{n}\sum_t^n \rho_\tau(e_t)M_t$ does not depend on η, h. Define

$$\begin{aligned}
\tilde{\xi}_{t1} &= (\tilde{q}_{\tau,t} - q_{\tau,t})\hat{f}_{-t}/f_t = \frac{1}{nh}\sum_{s\neq t}(q_{\tau,s} - q_{\tau,t})K_{st}/f_t, \\
\tilde{\xi}_{t2} &= \tilde{d}_t\hat{f}_{-t}/f_t = \frac{1}{nh}\sum_{s\neq t} b_s\epsilon_s K_{st}/f_t, \\
\tilde{\xi}_{t1} &= \check{b}_t(\eta - \eta_0),
\end{aligned}$$

where $\check{b}_t = \tilde{b}_t\hat{f}_{-t}/f_t = (nh)^{-1}\sum_{s\neq t} b_s K_{st}/f_t$. Then similar to the derivations in Racine and Li (2016), one can show that

$$\begin{aligned}
CV_{21}(\eta, h) &= \frac{1}{n}\sum_t^n [\tilde{\xi}_{t1} + \tilde{\xi}_{t2} + \tilde{\xi}_{\eta,t3}]v_t M_t \\
&= CV_{21,1}(h) + CV_{21,2}(h) + CV_{21,3}(\eta, h)
\end{aligned}$$

where $v_t = [1_{[e_t\leq 0]} - \tau]$ satisfies $E[v_t|X_t] = 0$.

Based on Lemma 4.2 and Lemma 4.3, $CV_{21,1}(h) = o_p\left((nh)^{-1}\right)$ and $CV_{21,2}(h) = o_p\left((nh)^{-1}\right)$. And $CV_{21,3} = O_p\left(n^{-\frac{1}{2}}|\eta - \eta_0|\right)$ because of $E[v_t|X_t] = 0$.

Therefore, we have

$$CV_{21} = O_p\left(n^{-1/2}|\eta - \eta_0|\right) + o_p\left((nh)^{-1}\right)$$

Following Racine and Li (2016), we have

$$\begin{aligned}
E[CV_{22}(\eta, h)|X_t] &\approx \frac{1}{2n}\sum_t^n M_t f_{e,t}[\tilde{\xi}_{t12}^2 + 2\tilde{\xi}_{t12}\tilde{\xi}_{\eta,t3} + \tilde{\xi}_{\eta,t3}^2] \\
&= CV_{22,1}(h) + CV_{22,2}(\eta, h) + CV_{22,3}(\eta, h)
\end{aligned}$$

where $f_{e,t} = f_e(0|X_t)$, $\tilde{\xi}_{t12} = \tilde{\xi}_{t1} + \tilde{\xi}_{t2}$.

Based on Lemma 4.4 to 4.6, we have

$$
\begin{aligned}
CV_{22,1} &= \frac{1}{2n}\sum_t M_t f_{e,t}\left[(\tilde{q}_{\tau,t}-q_{\tau,t})^2+2\tilde{d}_t(\tilde{q}_{\tau,t}-q_{\tau,t})+\tilde{d}_t^2\right]\hat{f}_{-t}^2/f_t^2 \\
&= CV_{22,1,1}+CV_{22,1,2}+CV_{22,1,3} \\
&= C_1h^4+\frac{C_2}{nh}+o_p\left((nh)^{-1}\right)
\end{aligned}
$$

where

$$
\tilde{q}_{\tau,t}=(nh)^{-1}\sum_{s\neq t} q_{\tau,s}K_{st}/\hat{f}_{-t},
$$

$$
\tilde{d}_t=(nh)^{-1}\sum_{s\neq t} d_sK_{st}/\hat{f}_{-t}
$$

and

$$
\hat{f}_{-t}=(nh)^{-1}\sum_{s\neq t}K_{st}.
$$

Based on Lemma 4.1 and Lemma 4.4, we have

$$
\begin{aligned}
CV_{22,2} &= \frac{1}{n}\sum_t[M_tf_{e,t}(\tilde{q}_{\tau,t}-q_{\tau,t}+\tilde{d}_t)\hat{f}_{-t}/f_t]\tilde{\xi}_{t,3} \\
&= B_{1,0}h^2(\eta-\eta_0)+O_p\left(n^{-1/2}|\eta-\eta_0|\right)
\end{aligned}
$$

where $B_{1,0}=E[M_tf_{e,t}B_{LC,t}b_t]$.

$$
CV_{22,3}=\frac{1}{2n}\sum_t M_tf_{e,t}\tilde{\xi}_{t,3}^2=B_{2,0}(\eta-\eta_0)^2
$$

where $B_{2,0}=(1/2)E[M_tf_{e,t}b_t^2]$.

Therefore, we have

$$
CV_{22}(\eta,h)=C_1h^4+\frac{C_2}{nh}+B_{1,0}h^2(\eta-\eta_0)+B_{2,0}(\eta-\eta_0)^2+O_p\left(n^{-1/2}|\eta-\eta_0|\right)+o_p\left((nh)^{-1}\right)
$$

$$\begin{aligned} CV(\eta,h) &= CV_{21}(\eta,h) + CV_{22}(\eta,h) \\ &= C_1 h^4 + \frac{C_2}{nh} + B_{1,0}h^2(\eta-\eta_0) + B_{2,0}(\eta-\eta_0)^2 \\ &\quad + O_p\left(n^{-1/2}|\eta-\eta_0|\right) + o_p\left((nh)^{-1}\right). \end{aligned}$$

Minimizing objective function $CV(\eta,h)$ with respect to η gives

$$\begin{aligned} \hat{\eta}-\eta_0 &= \frac{-B_{1,0}h^2}{2B_{2,0}} + O_p\left(n^{-1/2}\right) \\ &= -B_{11}h^2 + O_p\left(n^{-1/2}\right) \end{aligned}$$

where $B_{11} = E[M_t f_{e,t} B_{LC,t} b_t]/E[M_t f_{e,t} b_t^2]$.

Proof of Proposition 2.2

$$\begin{aligned} CV(\eta,h) &= C_1 h^4 + \frac{C_2}{nh} + B_{1,0}h^2(\eta-\eta_0) + B_{2,0}(\eta-\eta_0)^2 \\ &\quad + O_p\left(n^{-1/2}|\eta-\eta_0|\right) + o_p\left((nh)^{-1}\right) \\ &= C_3 h^4 + C_2 (nh)^{-1} + o_p\left((nh)^{-1}\right) \end{aligned}$$

the last equality is derived by substituting $\hat{\eta}-\eta_0 = -B_{11}h^2 + O_p\left(n^{-1/2}\right)$ into function $CV(\eta,h)$.

$$\begin{aligned} C_3 &= C_1 + B_{2,0}\left(\frac{B_{1,0}}{2B_{2,0}}\right)^2 - \frac{B_{1,0}^2}{2B_{2,0}} \\ &= C_1 - \frac{1}{4}\frac{B_{1,0}^2}{B_{2,0}} \\ &= \frac{1}{2}E\left[M_t f_{e,t} B_{LC,t}^2\right] - \frac{1}{2}\frac{\left(E[M_t f_{e,t} B_{LC,t} b_t]\right)^2}{E[M_t f_{e,t} b_t^2]}. \end{aligned}$$

Minimizing $CV(h)$ with respect to h gives

$$\hat{h} \approx (C_2/4C_3)^{1/5}\, n^{-1/5}$$

Proof of Theorem 2.1

$$\hat{q}_{\hat{\eta},\hat{b}}(x) - q_\tau(x) = \frac{\hat{m}(x)}{\hat{f}(x)},$$

where

$$\begin{aligned}
\hat{m}(x) &= (\hat{q}_{\hat{\eta},\hat{b}}(x) - q_\tau(x))\hat{f}(x) \\
&= \frac{1}{nh}\sum_t \left[a_t + \hat{b}_t\hat{\eta} + b_t\varepsilon_t - q_\tau(x)\right] K\left(\frac{X_t - x}{h}\right) \\
&= \frac{1}{nh}\sum_t \left[a_t + b_t\hat{\eta} - b_t\eta_0 + b_t\eta_0 + b_t\varepsilon_t - q_\tau(x)\right] K\left(\frac{X_t - x}{h}\right) + (s.o.) \\
&= \frac{1}{nh}\sum_t [q_\tau(X_t) - q_\tau(x)] K\left(\frac{X_t - x}{h}\right) \\
&\quad + \frac{1}{nh}\sum_t [b_t(\hat{\eta} - \eta_0)] K\left(\frac{X_t - x}{h}\right) + (s.o.) \\
&= \hat{m}_1(x) + \hat{m}_2(x) \\
&= A_1 + A_2 + A_3 + A_4
\end{aligned}$$

where $A_1 = E[\hat{m}_1(x)]$, $A_2 = \hat{m}_1(x) - E[\hat{m}_1(x)]$, $A_3 = E[\hat{m}_2(x)]$, $A_4 = \hat{m}_2(x) - E[\hat{m}_2(x)]$.

It is easy to show that

$$A_1 = f(x)B_{LC}(x)h^2 + o_p\left(h^2\right),$$

$$E[A_2^2] = \frac{1}{nh}\nu_0 b^2(x) f(x) + o\,(nh)^{-1}\,,$$

where $B_{LC}(x)$ is defined in Theorem 1 and $\nu_0 = \int K^2(v)dv$.

$$\begin{aligned} A_3 &= \frac{1}{nh}\sum_t E\left\{b_t(\hat{\eta} - \eta_0)K\left(\frac{X_t - x}{h}\right)\right\} \\ &= -B_{11}b(x)f(x)h^2 + o_p\left(h^2\right) \end{aligned}$$

where $B_{11} = \frac{E\left[M_t f_{e,t} B_{LC,t} b_t\right]}{E[M_t f_{e,t} b_t^2]}$.

$$A_4 = o_p\,(nh)^{-1}$$

Since $\hat{f}(x) = f(x) + o_p(1)$, therefore, we have

$$E[\hat{q}_{\hat{\eta},\hat{b}}(x) - q_\tau(x)] = \left[B_{LC}(x) - \frac{E[M_t f_{e,t} B_{LC,t} b_t]}{E[M_t f_{e,t} b_t^2]} b(x)\right] h^2 + o\left(h^2\right),$$

and

$$Var[\hat{q}_{\hat{\eta},\hat{b}}(x)] = \frac{1}{nh}\frac{\nu_0 b^2(x)}{f(x)} + o\,(nh)^{-1}\,.$$

Hence, we obtain

$$\sqrt{nh}[\hat{q}_{\hat{\eta},\hat{b}}(x) - q_\tau(x) - B_1 h^2] \xrightarrow{d} N(0, V(x))$$

where $B_1 = \left[B_{LC}(x) - \frac{E[M_t f_{e,t} B_{LC,t} b_t]}{E[M_t f_{e,t} b_t^2]} b(x)\right]$, and $V(x) = \frac{\nu_0 b^2(x)}{f(x)}$.

B.1 Appendix B

This Appendix presents some useful lemmas that are used for the proofs given in Appendix A.

Lemma 4.7. *Let $g_n(x_1, x_2)$ be a Borel function such that*

$$\int_{R_{2q}} |g_n(x_1, x_2)|^{1+\delta} dF_{1,1+j}(x_1, x_2) \leq M_n$$

and

$$\int_{R_{2q}} |g_n(x_1, x_2)|^{1+\delta} dF(x_1) dF(x_2) \leq M_n$$

for some $\delta > 0$ and $M_n > 0$. Then

$$\begin{aligned} & |\int_{R^{2q}} g_n(x_1, x_2) dF_{1,1+j}(x_1, x_2) - \int_{R^{2q}} g_n(x_1, x_2) dF(x_1) dF(x_2)| \\ \leq & \ 4M_n^{1/(1+\delta)} \beta^{\delta/(1+\delta)}(j) \end{aligned}$$

Proof: This is Lemma 1 of Yoshihara (1976).

Lemma 4.8. *Let $W_t \in R^r$ be a strictly stationary, β-mixing process with coefficient β_m. Let $\phi(\cdot, \cdot, \cdot)$ be a symmetric Borel measurable function defined on $R^r \times R^r \times R^r$ such that*

$$\begin{aligned} M_{n3} &= \max\{\int_{R^{3r}} |\phi(x_1, x_2, x_3)|^{1+\delta} dF_{i_1,i_2,i_3}(x_1, x_2, x_3), \int_{R^{3r}} |\phi(x_1, x_2, x_3)|^{1+\delta} dF(x_1) dF(x_2) dF(x_3)\} \\ M_{n12} &= \max\{\int_{R^{3r}} |\phi(x_1, x_2, x_3)|^{1+\delta} dF(x_1) dF_{i_2,i_3}(x_2, x_3), \int_{R^{3r}} |\phi(x_1, x_2, x_3)|^{1+\delta} dF(x_1) dF(x_2) dF(x_3)\} \end{aligned}$$

for all $1 \leq i_1 < i_2 < i_3 \leq n$.

Then

$$\begin{aligned} & E\left[\sum\sum\sum_{1\leq i_1<i_2<i_3\leq n} \phi(W_{i1}, W_{i2}, W_{i3})\right] \\ = & \ O(n^3 E[\phi(\tilde{W}_{i1}, \tilde{W}_{i2}, \tilde{W}_{i3})]) + O(n^2 M_{n12}^{1/(1+\delta)}) + O(n M_{n3}^{1/(1+\delta)}), \end{aligned}$$

where $\{\tilde{W}_t\}$ is an independent process that has the same marginal distribution as the depdendent process $\{W_t\}$.

Proof: This is Lemma 2 of Fan and Li (1999a).

Lemma 4.9. *Let $W_t \in R^r$ be defined as in Lemma 4.8 and let*

$$U_{n,k} = \binom{n}{k}^{-1} \sum_{1 \le t_1 < t_2 < ... < t_k \le n} H(W_{t_1}, ..., W_{i_k})$$

be a k-th order U-statistic. Define $H_1(w_1) = \int_{R^{r(k-1)}} H(W_1, ..., W_k) \prod_{i=2}^{k} dF(w_i)$ and $R_n = U_{n,k} - \theta - (k/n) \sum_{t=1}^{n} [H_1(W_t) - \theta]$, where $\theta = \int_{R^{rk}} H(W_1, ..., W_k) \prod_{i=1}^{k} dF(w_i)$. Further assume that $\beta_n^{\delta/(2+\delta)} = O(n^{-2+\epsilon})$ for some δ, $\epsilon > 0$. Then

$$R_n = O_p(n^{-1+\epsilon/2} s_\delta),$$

where $s_\delta = \sup_{1 \le t_1 < t_2 < ... < t_k \le n} (E|H(W_{t_1}, ..., W_{t_k})|^{2+\delta})^{1/(2+\delta)}$.

Proof: This is Proposition 2 of Denker and Keller (1983).